Measuring the Complexity of Musical Rhythm

Eric Thul

Master of Science

School of Computer Science

McGill University
Montreal, Quebec
June 2008

A thesis submitted to McGill University in partial fulfilment
of the requirements of the degree of Master of Science.

Acknowledgements

I would like to thank my past supervisor, now co-supervisor, Godfried T. Toussaint, for his valuable guidance, patience, and encouragement during my two years at McGill. I am grateful to Jörg Kienzle who graciously took on the responsibility of becoming my supervisor, as Godfried was donned Professor Emeritus. Also, I thank Francisco Gómez, who worked with Godfried and me on projects leading to this thesis. I owe much to these scholars who have helped me at every step, and have given me insight to academic research. Additionally, I would like to thank my parents, because they have given me so much over the years and have always supported my endeavors. Finally, I extend a warm thanks to Diane McClymont-Peace and Heather Peace, because the translation of my abstract into French would not have otherwise been possible. Also, Heather's support and kindness through my work on this thesis has been unfailing.

Abstract

This thesis studies measures of musical rhythm complexity. Informally, rhythm complexity may be thought of as the difficulty humans have performing a rhythm, listening to a rhythm, or recognizing its structure. The problem of understanding rhythm complexity has been studied in musicology and psychology, but there are approaches for its measurement from a variety of domains. This thesis aims to evaluate rhythm complexity measures based on how accurately they reflect human-based measures. Also, it aims to compare their performance using rhythms from Africa, India, and rhythms generated randomly. The results suggest that none of the measures accurately reflect the difficulty humans have performing or listening to rhythm; however, the measures do accurately reflect how humans recognize a rhythm's metrical structure. Additionally, the results suggest a need for normalization of the measures to account for variety among cultural rhythms.

Abrégé

Cette thèse examine les mesures de complexité rythme musicale. La complexité de rythme représente notre difficulté en jouant, en écoutant, ou en reconnaissant la structure, d'un rythme. La compréhension de la complexité étaít étudiée dans les champs de musicologie et de psychologie. Cependant, il y a d'autres méthodes de mesurage. Cette thèse a l'intention d'évaluer les mesures de complexité rythmique basé sur leurs exactitude de réfléchir leurs mesures humaines. Le but est aussi de faire comparer la performance des mesures en utilisant des rythmes de l'Afrique, de l'Inde, et des rythmes réalisés aléatoires. Les résultats suggèrent qu'aucun des mesures ne peuvent réfléchir exactement notre difficulté en jouant ou en écoutant des rythmes. Cependant, les mesures peuvent réfléchir avec précision notre moyen de reconnaitre la structure métrique de rythme. Les résultats suggèrent un besoin de normaliser les mesures pour rendre compte les variétés des rythmes culturels.

Table of Contents

Chapter 1

Introduction

Mathematics has been used to study music for quite some time. Indeed, since Pythagoras first discovered the relationship between numerical proportions and musical harmonies over 2500 years ago [133], the connection between mathematics and music has also been commented on by many other notable figures throughout history. For instance, the famous mathematician and philosopher, Gottfried Wilhelm Leibniz, indicated a deep, unconscious relationship between mathematics and music when he stated that, "music is the pleasure the human mind experiences from counting without being aware that it is counting." Henry David Thoreau, a renowned American author, elegantly pointed out a possible tie between the two domains with the words, "the most distinct and beautiful statement of any truth (as of music) must take at last the mathematical form." Additionally, the well-known American science and mathematics writer, Martin Gardner, observed that "a surprising proportion of mathematicians are accomplished musicians. Is it because music and mathematics share patterns that are beautiful?"

1.1 Context of the Problem

While statements recognizing the connection between mathematics and music are plentiful [7, 32], mathematical approaches to music have seen advances since the days of Pythagoras. In the 1930s, the composer and music theorist Joseph Schillinger developed a mathematically-based system for musical composition [138, 139, 140]. Schillinger devotes an entire book to aspects of music, such as rhythm, melody, and composition. He begins with a book on rhythm, where he touches upon principles of periodicity and synchronization, to generate new periodicities. Schillinger uses fundamental mathematical functions such as the least common multiple and greatest common divisor for grouping rhythmic structure [138]. Some viewed this work as a "cultural event of considerable import" [112], while others criticized it as an "example of misplaced ingenuity" [112]. Nevertheless, the Schillinger system sparked interest by the famous American composer George Gershwin who was a student of Schillinger for several years [112]. During this time, others in the domain of music applied mathematics to music theory. In the 1940s, the French composer Olivier Messiaen used mathematical properties of symmetry to explore musical rhythm [109]. However, even as mathematical principles began to gain popularity among music theorists and composers, there was one landmark event which strengthened the connection between mathematics and music: the advent of the digital computer. With this new technology, the foundations of mathematics could be used to analyze music in a new way.

In fact, the computer had been readily applied to the comparative analysis of music in 1949, by Bertrand Bronson from the University of California at Berkeley. He was the first to analyze music using the computer when he compared tunes of the British-American folk tradition [12]. Bronson coded elements from the musical pieces, such as melodic and rhythmic features, and compiled a large corpus to enhance his research with a multitude of ways for indexing the musical tunes for comparison [11, 12]. Work such as this lead researcher Tobias Robison, of Princeton University, to pose questions: "How can musical information be fed into a computer?" and "[W]hat sorts of things can a computer be made to do with this information?" [134] From questions such as these, many areas of music were approached; e.g., systematic methods for style analysis (La Rue's work [135] in 1962), the analysis of theory regarding atonal music (Forte's work [58] in 1965), and statistical analysis regarding components of music (Mendel and Lockwood's work [11] in 1967).

Since many computer-based music projects commonly share the need for *information* from the music, perhaps the words of Milton Babbitt in 1965, are most fitting: "the already vast research on information retrieval is probably the first object to which the music scholar should apply its techniques" [5]. However, perhaps Babbitt's words and the previously mentioned work of Mendel and Lockwood [11], as well as the work of others [12, 58, 84, 135, 142], were a few decades ahead of their time. It wasn't until about the turn of the millennium that techniques from the field of Information Retrieval had begun to be widely applied to music. The article by Bainbridge et al. [6] lit the fuse, so to speak, of the now booming field of *Music Information Retrieval* (MIR) [13]. J. Stephen Downie provides a definition for MIR: "a multidisciplinary research endeavor that strives to develop innovative content-based searching schemes, novel interfaces, and evolving networked delivery mechanisms in an effort to make the world's vast store of music accessible to all" [42]. Though, more recently, Tzanetakis et al. define MIR in a more general way: "[MIR is] an emerging interdisciplinary research area that encompasses *all aspects* of accessing digital music material (emphasis added)" [177].

As MIR is a widely interdisciplinary field, powerful approaches from Machine Learning, such as those found in Pattern Recognition, are being incorporated into the MIR domain. Underlying these approaches is the extraction of *features* [53, 62, 63, 108]. In general, a feature may be defined as, "a notable or characteristic part of something... something that helps distinguish one thing from another (or one group of things from another group of things)" [76]. Moreover, a feature must be chosen such that it is considered *salient*; i.e., evokes heightened attention [76]. The question now becomes: How does one know whether or not some feature evokes heightened attention with regard to music? One possible way to determine this, is to investigate what evokes attention from those who practice the schol-

arly study of music; i.e., musicologists. However, as pointed out by musicologist Richard Parncutt, "musicology today covers *all* disciplinary approaches to the study of *all* music in *all* its manifestations and *all* its contexts" [115]. Since the field of musicology has such breadth, as indicated by Parncutt's definition, then it admits a wide variety of topics which may be considered as a features of music.

1.2 Problem Statement

The feature of music considered in this thesis is *complexity*. Specifically, I am writing about formal measures of the complexity of musical rhythm. However, first consider the notion of complexity and how it can be applied to musical rhythm.

Depending on which academic field one considers, one may find a unique interpretation of what complexity means. For instance, the computer scientist's first thought may be that of Kolmogorov complexity, which defines complexity of a string of characters as the shortest program required to output (i.e., describe) that string [101]. The information theorist may first think of Shannon's information entropy, which describes complexity in terms of the shortest possible representation of a message [143]. If neither of these definitions come to mind, then perhaps one of the *forty-two* interpretations of complexity [103, 104], compiled by Seth Lloyd, would provide a suitable one. As Jeff Pressing states, "complexity is multi-faceted"; and thus, perhaps some combination of the forty-two definitions would best suit the broad notion that is complexity.

However, the application of such definitions of complexity, with respect to musical rhythm, is not new. Work by psychologist R. H. Stetson in 1905, mentions that a rhythm may be "more irregular, less 'pure' than that of the simple sound series" [154]. Thus, we see early experimental studies and theoretical work regarding rhythm, along with a notion of irregularity, which also may be considered to be an interpretation of complexity. Indeed, in a study by the psychologist H. E. Weaver in 1939, he describes such rhythm irregularity to be characteristic of musical *syncopation* [182]. Syncopation will be covered later; however, consider it to be another (musically oriented) form of complexity [106, 125].

In addition to such early studies, research in the 1960s began to apply Shannon's information entropy to determine the complexity of rhythms. This is exemplified by the work of Paul Vitz [179] in 1968, and one year later by Vitz and Todd [180] in 1969. The work that followed these studies on rhythms presented a wide range of techniques and definitions regarding the measurement of rhythm complexity. From the psychological literature, studies comparing rhythm complexity measures to subjective human-based measures have been presented by Povel and Essens [119], in 1985, Peter Essens [48] in 1995, Shmulevich and Povel [146] in 2000, and most recently, Fitch and Rosenfeld [57] in 2007. Furthermore, measures of rhythm complexity have been seen in artificial intelligence [157, 158],

cognitive psychology [106, 125, 126, 150], computational geometry and computer science [40, 67, 68, 86, 166, 173], engineering [145, 147], and psychology [48, 114, 121, 146].

Clearly, Pressing's statement about complexity being multi-faceted rings true, as can be seen by the multitude of approaches in the literature regarding musical rhythm. Thus, herein lies the problem: *how does one accurately, adequately, and robustly measure the complexity of musical rhythm?* This thesis studies formal measures of the complexity of musical rhythm. We want to determine, out of the vast techniques, how to appropriately quantify the complexity of musical rhythm. The rationale behind this study is to further understand which techniques most accurately reflect the way humans interpret aspects inherent to the complexity of musical rhythm. There are three main goals in this thesis.

1. Implement and validate formal rhythm complexity measures in the literature.
2. Evaluate rhythm complexity measures regarding human-based measures of rhythm complexity gathered from previous psychological experiments; i.e., *human perceptual complexity, human performance complexity*, and *human metrical complexity*.
3. Demonstrate the performance of the rhythm complexity measures using rhythms found in African and Indian cultures, and also using randomly generated rhythms.

1.3 Contribution

Each of the main goals presents a significant contribution to work regarding the computational techniques of measuring musical rhythm complexity. First, implementing and evaluating a wide range of rhythm complexity measures, is important because, while most of the measures have previously been implemented, there is a handful which have not. Moreover, one important measure by Longuet-Higgins and Lee [106] was discovered by this thesis to be previously implemented [57, 150] incorrectly with respect to a rule specified by Longuet-Higgins and Lee. Thus a correct version is presented and discussed here. In addition, many of the measures have been expanded and implemented to handle a wide-variety of rhythms. But perhaps most importantly, since each of the complexity measures has been implemented for this thesis, this is the first known instance where they have all been compiled into one study, and also standardized and integrated into a wrapper program for convenient use and testing. Those in the MIR field may find this code beneficial for feature extraction tools or those for rhythm analysis, such as query by rhythm tools [18].

Second, evaluating the rhythm complexity measures regarding three types of human-based measures of rhythm complexity, is perhaps most significant to musicologists and ethnomusicologists who study the complexity of music in regional contexts [47, 149]. Such computational tools may, as Tzanetakis et al. remark, "provide the potential to assist" [177], in their definition of Computational Ethnomusicology by allowing for large data sets to be analyzed in little time, in a variety of ways. This thesis brings together the previous evalu-

ations of Povel and Essens [119], Essens [48], and Shmulevich and Povel [146], regarding human-based rhythm complexity measures. Additionally, this thesis incorporates the study by Fitch and Rosenfeld [57], using their valuable data to further extend the previous evaluations of rhythm complexity measures. This is significant because no other known work has presented a standard method for evaluating the numerous complexity measures in the literature, and thus there has been no basis for direct comparison. Smith and Honing's [150] work is an exception since they compared, by linearly scaling, the complexity values from a small subset of complexity measures. However, the exact method for scaling was not discussed in their work, and the measure they implemented and declared to perform best, was found by this thesis not to follow the algorithm's original definition. Another exception is previous work by Gómez, Thul, and Toussaint [68], expanded upon here.

Third, demonstrating the performance of the complexity measures on rhythms found in African and Indian cultures and on randomly generated rhythms, is significant because this is the first know experiment to do so. Previous experimentation has been on rhythms which were synthetically generated, as is the case with Povel and Essens' data [119], Essens' data [48], and Shmulevich and Povel's data [146]. In the case of Fitch and Rosenfeld's data, they generated their rhythms without regard to those found in African and Indian culture; however, a few that they generated did correspond to rhythms from those cultures. Regardless, they did not compare measures of complexity in their study. Hence, here we present rhythms from two cultures and demonstrate how each measure performs. One exception is previous work by Thul and Toussaint [164], expanded on here.

The main result of this thesis is that none of the approaches for measuring rhythmic complexity, accurately measured the complexity of musical rhythm on the whole. Instead, the measures seem to most-closely reflect the human-based measure of metrical complexity. Moreover, this thesis has shown that when using rhythms found in African and Indian cultures, there is a need for normalization to handle the varied rhythms found in such cultures. Additionally, improvements regarding some of the measures are proposed in light of the diverse rhythms found in the experimental data of this thesis.

However, before such conclusions are drawn, Chapter 2 presents a preliminary discussion on definitions and notation. Chapter 3 presents the rhythm complexity measures, providing a description of each. Chapter 4 describes the rhythms used to evaluate the complexity measures, and discusses human-based measures of rhythm complexity. Chapter 5 develops the methodology for evaluating the measures. Chapter 6 presents the results from our methods, and Chapter 7 provides a discussion of the results, and concludes the thesis.

Chapter 2

Preliminaries

Consider the following discussion which includes definitions, notations, and conventions to be held throughout the rest of this thesis. First, we will define musical terms used in this thesis: *rhythm*, *pulse*, *onset*, *meter*, *beat*, and *time signature*. Second, we will review the different representations of musical rhythm and meter used in the literature. Third, we will define the notion of musical syncopation, which was mentioned in Chapter 1.

2.1 Definitions

Musical rhythm has alluded to precise definition for many years. In fact, the well-known musicologist and author, Curt Sachs, states that rhythm is "a word without a generally accepted meaning" [136]. For the computer scientist, this is bad news, because specific definitions are generally a necessity when computationally modeling a phenomenon such as musical rhythm. Thus, the working definition of *rhythm* used in this thesis is derived from an article in *Grove Music Online* by Justin London [105]. We define *rhythm* to be the description and understanding of the duration and durational patternings of musical notes.

In our definition of rhythm, the term *duration* is presented. A duration is the number of time units between musical notes. In this thesis, we take the fundamental (and indivisible) time unit to be the *pulse*. Consider the following definition of *pulse* derived from Cooper and Meyer. We define a *pulse* to be "one of a series of regularly recurring, precisely equivalent stimuli" [29].

A pulse is similar to a pulsation generated by a metronome, and so a duration is composed of some number of pulses. However, in order to mark where the durations begin, we consider marked time units (pulses) of instantaneous attack; i.e., striking a drum or sounding a musical note. This instantaneous attack is called an *onset*. The definition used in this thesis is also derived from Cooper and Meyer [29]. We define an *onset* to be an instantaneous attack which marks the beginning of a duration of pulses. Referring back to our definition of rhythm, the onsets create the durational patternings, which are measured in terms of the pulses contained in each duration.

Now consider a definitions of *meter* and *beat*. These definition are derived from London [105]. We define *meter* to be the equal subdivision of the pulse, marked by strongly accented pulses, which we define as *beats*. Thus, meter represents a structure of the pulses, where this structure is realized by the beats. We define a *time signature* as notation which suggests the structure of the meter; i.e., where the beats are. In this thesis we use time signatures $4/4$ to indicate 4 beats in 16 pulses, $2/4$ to indicate 2 beats in 16 pulses, $3/4$

to indicate 3 beats in 12 pulses, and $6/8$ to indicate 6 beats in 12 pulses [114]. These definitions, as previously discussed, shall be used throughout the remainder of this thesis.

2.2 Representations

Consider the following ways in which a rhythm may be represented. There are five main methods discussed below. The first two show musically oriented notation for displaying rhythm. Then we discuss two methods which are popular ways to describe the skeletal pattern of the rhythm; i.e., removing musical aspects except the pattern of durational values. Finally, the last representation discussed shows how rhythm may be visualized geometrically.

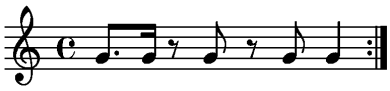


Figure 2.1: *The* clave son *rhythm in music notation.*

The first method shown in Figure 2.1 is perhaps the most familiar to the musician. This shows the famous Afro-Cuban rhythm, the *clave son*. The rhythm is presented in one repeating measure of $4/4$, or common time, thus, we have a total of 16 sixteenth-notes in the measure. Throughout this thesis, rhythms are assumed to be repeated indefinitely, unless otherwise specified, and the smallest music note value considered is the sixteenth note. Both of these assumptions are due to the nature of the rhythm data presented in Chapter 4. Taking such assumptions into account, consider the next representation.

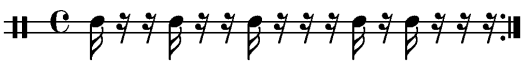


Figure 2.2: *The* clave son *rhythm in percussion music notation.*

Figure 2.3 shows notation for striking a percussion instrument. Notice that the smallest note is a sixteenth note, and that there is a mixture of sounded notes and silent notes (i.e., rests), totaling 16 pulses. Thus, by the definition of pulse, we have the sixteenth note representing the smallest indivisible time unit in each example of music notation. As will be seen, not all rhythms will have 16 pulses. The value of the pulse however, will remain constant. We wish to remove all music notation from a rhythm and only present the duration pattern. Thus, we consider a representation more fitting to the requirements of this thesis.

Figure 2.3 shows an example of the *clave son* rhythm in *Time Unit Box System (TUBS)* notation, here-after referred to as *box notation*. Box notation was first used by musicologists Philip Harland [166] and James Koetting [90] when studying African polyrhythms. This

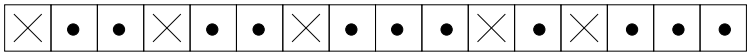


Figure 2.3: *The* clave son *rhythm in box notation (Time Unit Box System).*

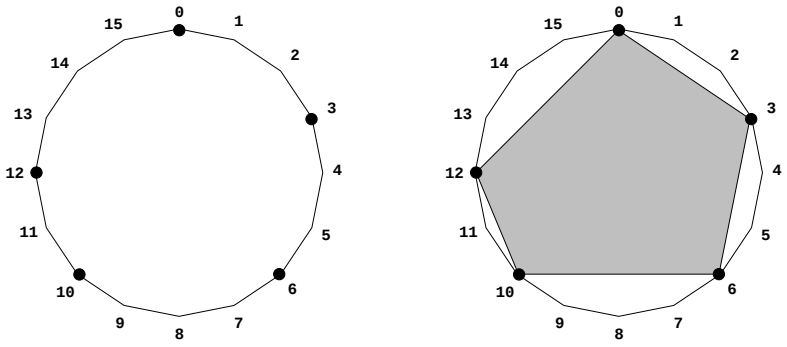
notation removes the traditional music notation from the rhythm and clearly shows the durational pattern, which nicely fits with the definition of rhythm in this thesis. The onsets are marked by an "x" and the silent pulses are marked by a ".". Thus, box notation shall be the standard when discussing rhythms.

1 0 0 1 0 0 1 0 0 0 1 0 1 0 0 0

Figure 2.4: *The* clave son *rhythm in binary notation.*

Figure 2.4 shows the *clave son* rhythm as a binary string, presenting another notation which describes duration patterns. This notation is used in the domain of computer science [166], and some complexity measures in Chapter 3 will rely on a binary representation. However, one may notice that this is the same as box notation, except the onsets are marked by a "1" and the silent pulses are marked by a "0". This notation shall be interchangeably used with box notation when needed.

Binary notation, along with box notation are skeletal representations of the rhythm which retain purely durational patterns. These are ideal for this thesis. However, from such patterns, we may also consider a more visual approach to representing rhythms.



(a) Points around the circle. (b) An inscribed polygon.

Figure 2.5: *Geometric representations of the* clave son *rhythm, where (a) shows the 16 pulses evenly distributed around the circle with black dots marking the points corresponding to an onset. In (b), adjacent black dots are connected by straight line segments inscribing a polygon inside the circle.*

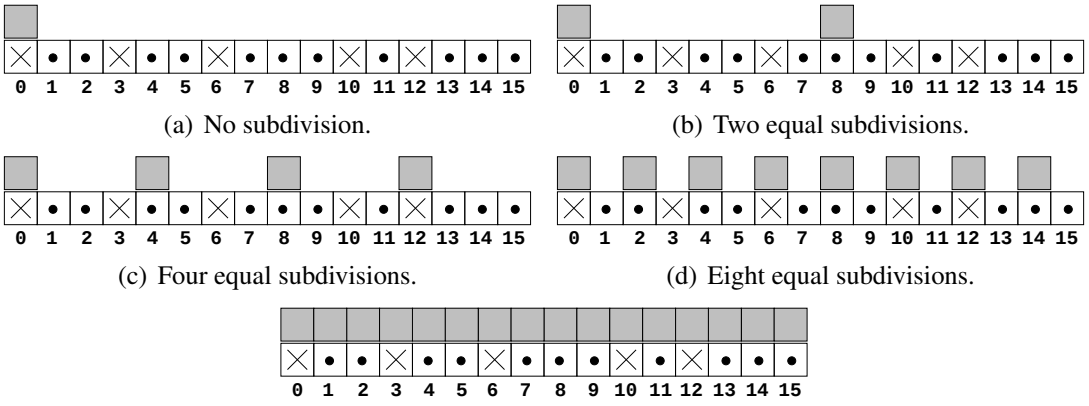
In Figure 2.5, there are two examples of ways to display a rhythm geometrically [172]. Figure 2.5(a) shows the *clave son* rhythm placed around a circle starting at zero (i.e., twelve)

o'clock. Each numbered point represents a pulse of the rhythm, where the numbered points marked with a black dot indicating an onset. Essentially one places a rhythm on an equally-spaced subdivision of points around the perimeter of a circle, where the number of points corresponds to the number of pulses. Figure 2.5(b) takes this one step further by inscribing a polygon inside the circle. Such a polygon is inscribed by simply connecting the adjacent marked points (black dots; onsets) from the previous example in Figure 2.5(a). Such geometric representations have been seen before, notably in the work of Godfried Toussaint [166, 168, 172]. Thus, we present such geometric forms since we include geometric techniques to measure complexity in Chapter 3.

2.3 Syncopation

The notion of syncopation has various interpretations by music scholars [77]. The working definition chosen for this thesis has been selected to accord with the previous definitions regarding rhythm. In *The Harvard Dictionary of Music*, *syncopation* is defined as "a momentary contradiction of the prevailing meter or pulse" [130].

Recall that the definition of meter used here is an equal subdivision of the pulse, and the equal subdivisions are marked by beats. Now, there are many ways to subdivide the pulse into equal parts; e.g., when there are 16 pulses, the equal subdivisions groups of: (a) 1, which means no dividing occurs. (b) 2, (c) 4, (d) 8, or (e) 16. In this example of 16 pulses, each equal subdivision is a power of two. To visualize this, consider Figure 2.6, which shows where the beats are for each metrical level of a rhythm with 16 pulses.



(a) No subdivision. (b) Two equal subdivisions.

(c) Four equal subdivisions. (d) Eight equal subdivisions.

(e) Sixteen equal subdivisions.

Figure 2.6: *The metrical levels of a rhythm with 16 pulses, where the pulses are numbered 0 through 15. At each level, the pulses considered to be a beat are marked above with a grey square.*

For each level shown in Figure 2.6, if we count the number of times each pulse is marked as a beat, and place that number of grey squares on the corresponding pulse, we arrive at

Figure 2.7 below. For example, pulse 0 is marked as a beat in all five case, and thus has five grey squares.

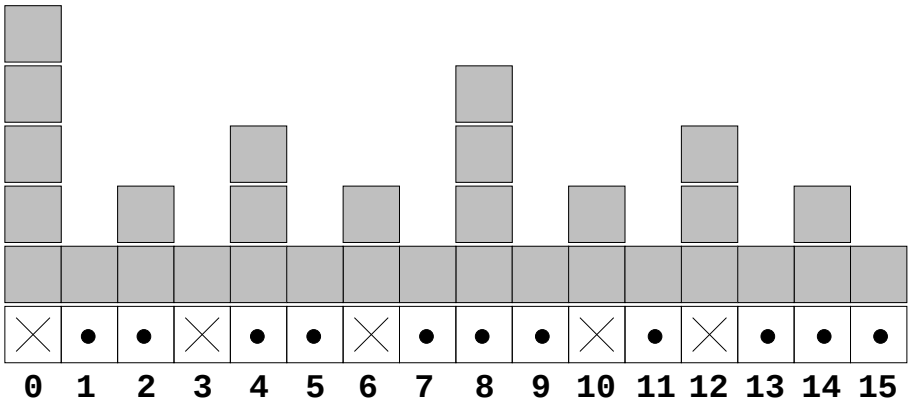


Figure 2.7: *The metrical hierarchy with corresponding weights of a 16-pulse rhythm. The weights are determined by joining all the metrical levels.*

The importance of Figure 2.7 is that it represents a metrical hierarchy, which was established in the authoritative text, *A Generative Theory of Tonal Music*, by Lerdahl and Jackendoff [99]. Also, this hierarchical model has been seen in work by Maury Yeston [185]. Regarding the metrical hierarchy, each pulse receives a metrical weight (the number of grey squares above the pulse). When an onset is sounded at a pulse with a high weight, this is considered to agree with the prevailing meter. However, when an onset is sounded at a pulse with a low weight, the prevailing meter is contradicted, and therefore we consider this onset to be syncopated [99, 161]. As an example, in Figure 2.7, the onset at pulse 3 is considered more syncopated than at pulse 12.

This notion of syncopation in terms of going against the metrically salient pulses, is the one adopted by this thesis. This is because some of the measures to be discussed in Chapter 3 will rely on such a structure to measure syncopation, which is considered here to be a form of rhythmic complexity in a metrical framework. Hence, this definition of syncopation shall be applied, along with the other definitions associated with rhythm in the complexity measures to be discussed.

Chapter 3

Complexity Measures

In the following we review the rhythm complexity measures evaluated in this thesis. There are six main categories into which the measures may be broken down. First, we present *rhythm syncopation* measures that use metrical weights determined from the metrical hierarchy to find syncopated onsets in a rhythm. Second, we discuss *pattern matching* measures which chop up a rhythm as determined by the levels of the metrical hierarchy and search each sub-rhythm for patterns which indicate the complexity. Third, *distance* measures are covered, which measure how far away a rhythm is from a more simple rhythm composed of beats. Fourth, we discuss purely mathematical measures which are based upon Shannon's *information entropy* [143]. Fifth, measures which pertain to the shape of histograms, which are generated from a rhythm's durational patterns, are presented. We call these, *interonset interval histogram* measures. Finally, measures of *mathematical irregularity* are discussed, which measure geometric properties of the rhythm when placed on a circular lattice.

3.1 Rhythmic Syncopation

The two complexity measures described in this section pertain to the metrical hierarchy of Lerdahl and Jackendoff [99], which was previously mentioned in Chapter 2. Thus, the measures below utilize a weighting system on this hierarchy and measure how syncopated a rhythm is, in order to measure the complexity. Consider the following measures presented by Toussaint [166] and Longuet-Higgins and Lee [106].

3.1.1 Toussaint's Metrical Complexity

The *Metrical Complexity* measure proposed by Toussaint [68, 166] calculates a sum of the metric weights for each onset in a given rhythm. Thus, the Metrical Complexity builds on the previous definition of syncopation seen in § 2.3. Recall the image of the weighted metrical hierarchy, reproduced below in Figure 3.1.

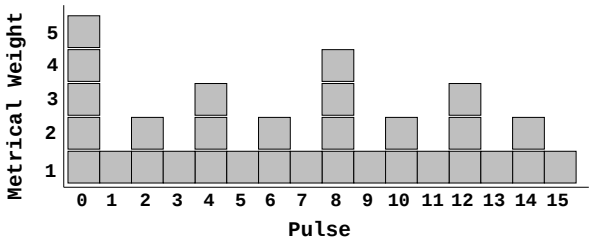


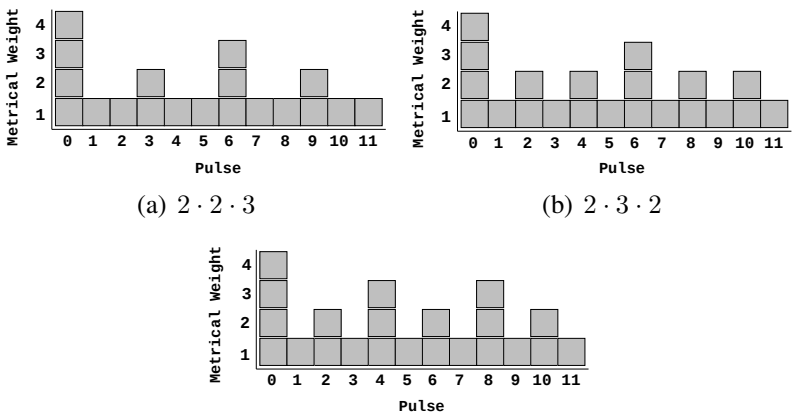
Figure 3.1: *The weighted metrical hierarchy for rhythms with 16 pulses.*

In Figure 3.1, we have the pulses along the x-axis, and the corresponding metrical weight on the y-axis, so for any pulse, we know how metrically important it is in the hierarchy. For example, the pulse at position 0 is the most important, followed by the pulse at position 8. Let us now consider a general construction for this hierarchy described in the five-step process below. This process is adapted from Toussaint's description [166].

First, consider the number of pulses of a rhythm to be n. Find all prime numbers which divide n, ignoring 1. Essentially one creates a prime factorization of n; e.g., when $n = 16$ we have a prime factorization of $2 \cdot 2 \cdot 2 \cdot 2$. Each unique permutation of the list of prime factors will yield a distinct metrical hierarchy. Second, pick one permutation. When $n = 16$ we only have one choice. Let ℓ be an integer which corresponds to the current metrical level, initialize $\ell = 1$. Define a vector $\mathbf{w}$ to keep track of the metrical weight for each pulse, index $\mathbf{w}$ by $0 \leq i \leq 15$, setting each w_i to zero. Third, loop through $\mathbf{w}$ by i, from $i = 0$ to 15 (inclusive), and add one to each weight w_i where n/ℓ is the step-size of the increment in the loop. Hence, since ℓ is 1, we increment i by $n/\ell = 16$ at each iteration, adding one to only index 0 of $\mathbf{w}$. Fourth, set ℓ to be the product of ℓ multiplied by the first prime factor in the prime factorization of n, then discard that prime factor. Here we would multiply ℓ by 2. Fifth, repeat at step three with the new value ℓ. Stop when there are no more prime factors to multiply by. In the case of $n = 16$, we would have ℓ equal to 1, 2, 4, 8, and 16. Thus, the step-size in the loop at step three would be 16, 8, 4, 2, and 1.

From the process above, one generates the hierarchy in Figure 3.1 when $n = 16$. However, in this case $n = 16$ conveniently admits only one hierarchy, since its prime factorization is $2 \cdot 2 \cdot 2 \cdot 2$. Since the process is generally stated, consider the case for $n = 12$. Here, there are three unique permutations of its prime factorization: (a) $2 \cdot 2 \cdot 3$, (b) $2 \cdot 3 \cdot 2$, and (c) $3 \cdot 2 \cdot 2$. Thus we perform the five step process above and generate three metrical hierarchies. Consider Figure 3.2, which depicts each case.

Using the weighted metrical hierarchy, the Metrical Complexity measure is calculated by looking up the weight of each onset in a rhythm. So, for each onset in a rhythm, take its pulse index, and find the weight that pulse has in the corresponding metrical hierarchy. The Metrical Complexity is the sum of the corresponding weights. For example, the *clave son*, [x . . x . . x . . . x . x . . .] has onsets at positions: $(0, 3, 6, 10, 12)$, corresponding to metrical weights: $(5, 1, 2, 2, 3)$. The Metrical Complexity is $5 + 1 + 2 + 2 + 3 = 13$. However, since the pulses with a higher weight are those which are less syncopated, a smaller value means the rhythm is more complex. Moreover, when we have more than one metrical hierarchy to pick from, as in the case when $n = 12$, we find the Metrical Complexity using each hierarchy and then take the average to be the final Metrical Complexity. This measure is abbreviated as *metrical*.



(a) $2 \cdot 2 \cdot 3$ (b) $2 \cdot 3 \cdot 2$

(c) $3 \cdot 2 \cdot 2$

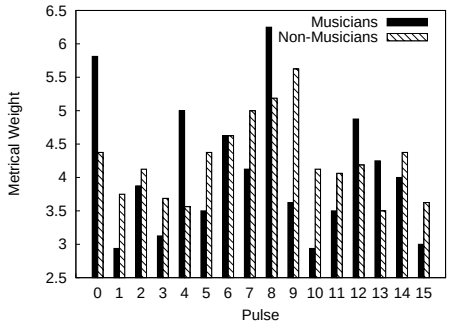
Figure 3.2: *The weighted metrical hierarchies for rhythms with 12 pulses.*

The Metrical Complexity measure has been investigated further by applying two different weighting schemes, rather than the one corresponding to the metrical levels, as described above. Moreover, we present three types of normalization. Below, we first consider each new weighting scheme, and then consider each normalization method for comparing rhythms with a different number of pulses and onsets.

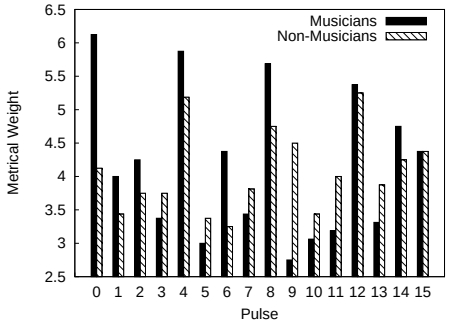
Palmer and Krumhansl

Thus far the metrical weighting scheme used in the Metrical Complexity measure has been the one proposed by Lerdahl and Jackendoff [99], in 1983. However, one might question whether or not people listen according to the proposed weighting scheme. In fact, Palmer and Krumhansl [114] did ask such a question. Their aim was "to determine whether or not impoverished temporal contexts evoke hierarchically structured knowledge about metrical structure, even when this structure is not contained in the sensory context" [114]. An experiment was devised to study musicians' and non-musicians' knowledge of meter. In other words, Palmer and Krumhansl tested how each group empirically weighted the pulses of metrical hierarchy (see Palmer and Krumhansl's Figure 3 [114] and Table 2 [114]).

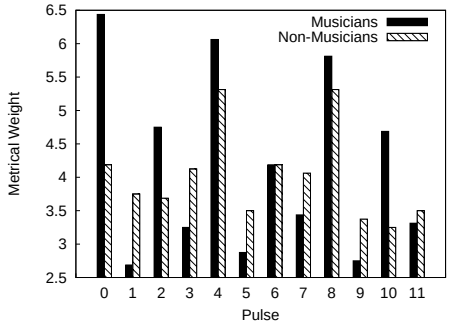
The empirical data from Palmer and Krumhansl's study on metrical hierarchies [114] has been applied to the Metrical Complexity measure. Smith and Honing [150] also utilized the results from Palmer and Krumhansl. However, in this context, we apply the empirical weights from the group of musicians and that of non-musicians for rhythms with 16 pulses and 12 pulses. Moreover, in each case we have two sets of weights: those regarding $4/4$ and $2/4$ time signatures for 16 pulse rhythms, and those regarding $3/4$ and $6/8$ time signatures for 12 pulse rhythms. Consider Figure 3.3 which shows graphs for the weights of each pulse using the different empirical results.



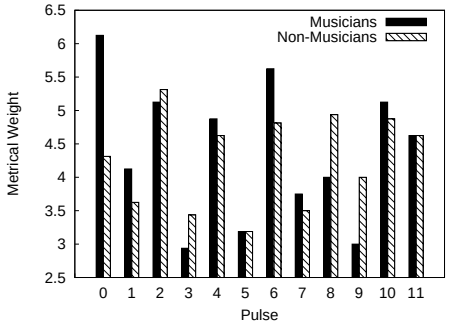
(a) Time signature $2/4$.



(b) Time signature $4/4$.



(c) Time signature $3/4$.



(d) Time signature $6/8$.

Figure 3.3: *The weighted metrical hierarchies for rhythms with 12 pulses and 16 pulses according to Palmer and Krumhansl's study.*

Applying each weighting scheme, we arrive at eight measures, abbreviated *metricalpk-MUS-24, metricalpk-NMUS-24, metricalpk-MUS-44, metricalpk-NMUS-44, metricalpk-MUS-34, metricalpk-NMUS-34, metricalpk-MUS-68*, and *metricalpk-NMUS-68*.

Euler

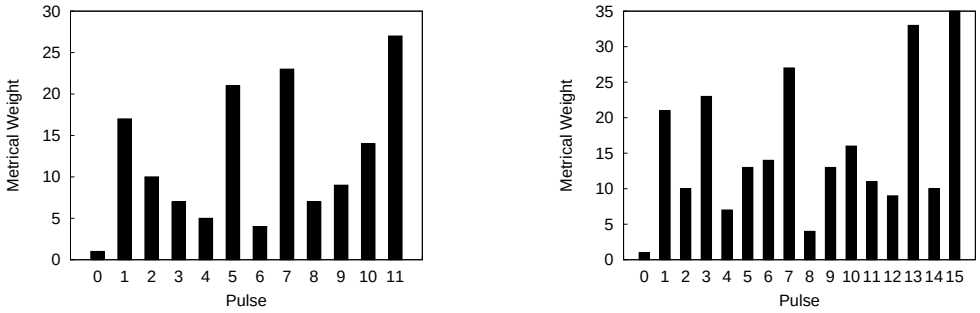
A second weighting scheme proposed by Gonzalez [69] uses what he calls the *Euler Complexity* of a number to determine the weight of each pulse in a rhythm. The procedure for calculating the Euler Complexity proceeds as follows, this procedure is adapted from Gonzalez's work [69].

Let n be the number of pulses. For $0 \leq i \leq n-1$, we will determine its Euler Complexity. At each iteration we must calculate the following values. First we determine the ratio between *least common multiple*, denoted lcm, of $i+n$ and n, and the *greatest common denominator*, denoted gcd, of $i+n$ and n. We denoted this ratio by m.

$$m = \frac{lcm(i+n, n)}{gcd(i+n, n)} \tag{3.1}$$

Next we determine the prime factorization of m, and apply Euler's formula for the complexity of m using the prime factorization. Let $c(m)$ be the complexity, $\mathbf{p}$ be a vector of unique prime factors, and $\mathbf{q}$ be the corresponding vector of the number of times each prime factor appears in the factorization. Consider Equation (3.2)

$$c(m) = 1 + \sum_{p \in \mathbf{p}, q \in \mathbf{q}} (p-1) \cdot q \tag{3.2}$$



(a) Weights for 12-pulse rhythms.

(b) Weights for 16-pulse rhythms.

Figure 3.4: *Euler metrical weights for rhythm with 12 pulses and 16 pulses, as described by Gonzalez [69]. Note that a higher value in this case means the rhythm is considered to be more complex.*

Thus, we have two new weighting schemes which can be applied to the Metrical Complexity measure. Consider Figure 3.4, which shows the resulting weights for rhythms with

12 and 16 pulses. The *Metrical Complexity (Euler Weighted)* measure incorporating these schemes is abbreviated: *metricaleuler*. Note that here, a higher value is more complex.

Onset Normalization

In addition to alternate weighting schemes, a normalization method is tested for the Metrical Complexity measure, which was presented by Toussaint [166] in 2002. The basic idea is to account for rhythms with a different number of pulses and onsets, thus making the measure relative. To do this, there is one additional step when calculating the complexity. Once we find the Metrical Complexity of a rhythm, we then look back at the corresponding metrical hierarchy. Give that our rhythm has k onsets, we pick the k largest metrical weights and sum them together. We then subtract the Metrical Complexity value from this sum. This indicates how far away the rhythm is from one which has its k onsets placed on pulses with the highest weight. We call this *Metrical Complexity (Onset Normalized)*, and abbreviate it *metricalonorm*. Also, for the weighting schemes, we have *metricalonormpk-MUS-24, metricalonormpk-NMUS-24*, *metricalonormpk-MUS-44, metricalonormpk-NMUS-44, metricalonormpk-MUS-34*, *metricalonormpk-NMUS-34, metricalonormpk-MUS-68*, *metricalonormpk-NMUS-68*, and *metricalonormeuler*.

Pulse Normalization

The next form of normalization accounts for rhythms with a different number of pulses. Assume that we calculate the Metrical Complexity as usual. Then, to normalize, we divide the complexity value by the sum of all the metrical weights in the hierarchy. So, it is as if one scales the Metrical Complexity of a rhythm with n pulses by the Metrical Complexity of an imaginary rhythm with n pulses, which has an onset on every pulse. The *Metrical Complexity (Pulse Normalized)* measure is abbreviated *metricalpnorm*, and for the weightings we have, *metricalpnormpk-MUS-24, metricalpnormpk-NMUS-24, metricalpnormpk-MUS-44, metricalpnormpk-NMUS-44*, *metricalpnormpk-MUS-34*, *metricalpnormpk-NMUS-34*, *metricalpnormpk-MUS-68*, *metricalpnormpk-NMUS-68*, and *metricalpnormeuler*.

Pulse-Onset Normalization

The last normalization uses the result from the Metrical Complexity (Pulse Normalized) measure and simply scales that value by the number of onsets in the rhythm. This normalization is termed the *Metrical Complexity (Pulse-Onset Normalized)* measure with the abbreviation being *metricalponorm*. Concerning each of the different weightings schemes, we have *metricalponormpk-MUS-24, metricalponormpk-NMUS-24*, *metricalponormpk-MUS-44, metricalponormpk-NMUS-44, metricalponormpk-MUS-34*, *metricalponormpk-NMUS-34*, *metricalponormpk-MUS-68*, *metricalponormpk-NMUS-68*, and *metricalponormeuler*.

3.1.2 Longuet-Higgins and Lee Complexity

Longuet-Higgins and Lee [106] in 1984 proposed a context-free grammar for musical rhythm, similar to the grammars used to capture the syntactic structure of sentences. They define a rule-set to, "designate 'metrical units' at various levels in a metrical hierarchy" [106].

We can describe the model of Longuet-Higgins and Lee by the familiar weighted metrical hierarchy previously described in §§ 3.1.1. The idea behind generating Longuet-Higgins' and Lee's model is exactly the same, except it is formulated in terms of a tree structure where the terminal nodes retain the metrical weight for each pulse. Let us describe the process of generating such a tree.

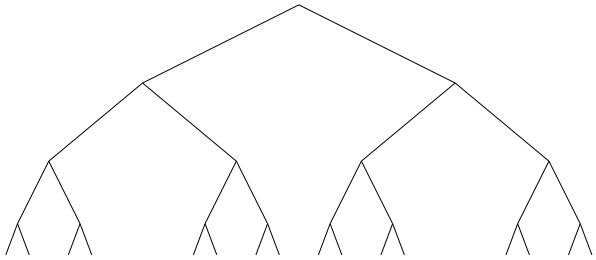
First, let n be the number of pulses of the rhythm in which we shall generate it's metrical hierarchy. Second, find the prime factorization of n where we ignore 1 (as we've done in the last section). Keep the prime factors in a list. Again, of the unique permutations of the prime factors in the list, pick one. Third, generate a tree by iterating over the selected permutation of the list of prime factors. We start with a single node in the tree, called the root. Then, at each iteration, we perform the following, where p is the current prime factor in the list and ℓ (initialized to 1) is the current level (depth) in the tree.

1. For all the nodes m at level ℓ, create p children whose common parent is m.
2. Increment ℓ by 1.
3. Check p off the list of prime factors, and pick the next prime factor in the list.

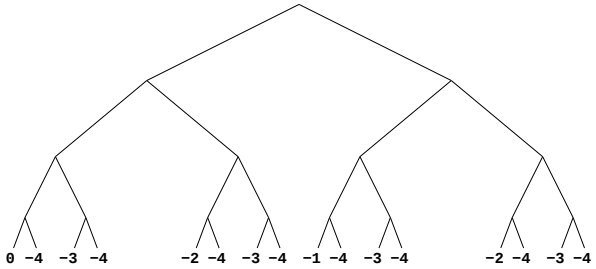
Following this process will generate a tree, and in the case of $n = 16$, we have the prime factor list being $2{\cdot}2{\cdot}2{\cdot}2$, and so a binary tree with 5 levels is formed, shown in Figure 3.5(a).

The next series of steps determines the weight on each terminal node of the tree, such as of the tree in Figure 3.5(a). We do this to determine the metrical weight of each pulse of rhythm. To do so, we may again draw upon the process from §§ 3.1.1. Thus, in step four, we recall the prime factorization list used to generate the tree. Again we iterate over each prime factor. This time define ℓ to be metrical level and initialize it to 1. Also, define the vector $\mathbf{w}$ to keep track of the weight of each pulse (i.e., terminal node), where all the weights are initialized to zero. Note that scanning the terminal nodes from left to right in the tree corresponds to indices 0 through 15 in $\mathbf{w}$.

In step five, loop through $\mathbf{w}$ for $0 \leq i \leq 15$. For each w_i, we always subtract 1, unless the i is 0 or i is a multiple of (n/ℓ). After the loop, set ℓ to be the product of ℓ multiplied by the next prime factor in the list; cross off that prime factor. For step six, we repeat step five with the new value of ℓ. We stop when there are no more prime factors by which to multiply ℓ. For example, when we have our prime factor list for $n = 16$; i.e., $2 \cdot 2 \cdot 2 \cdot 2$, we have ℓ taking values 2, 4, 8, and 16 upon each repeat of step five. This means that we



(a) Metrical hierarchy.



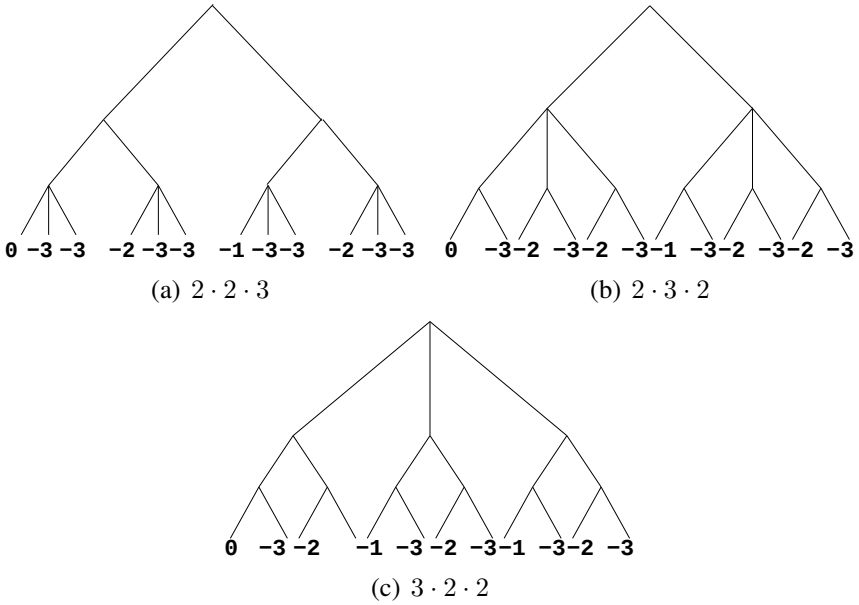
(b) Weighted metrical hierarchy.

Figure 3.5: *Binary tree representing the Longuet-Higgins and Lee metrical hierarchy structure for a rhythm with 16 pulses both unweighted and weighted.*

do not subtract a 1 from the weight vector when the index is 0 or a multiple of 8, 4, 2, and 1, respectively, each time we repeat the loop process. Consider Figure 3.5(b), which depicts the weights on the metrical hierarchy for $n = 16$. We also handle the case when there is more than one unique permutation of the prime factorization of n, such as when $n = 12$. Again, we may use the described procedure to generate different weighted metrical hierarchies. Consider the three hierarchies for $n = 12$ in Figure 3.6 below.

Once the metrical hierarchy has been generated, the complexity of the rhythm is calculated by looking for silent pulses which have a higher weight than the onset which precedes them [106]. Thus, for a rhythm in box notation with n pulses and with a weight vector $\mathbf{w}$, which carries a metrical weight for each pulse as described above, we may measure the complexity as follows. First loop through each pulse of the rhythm, indexed by $0 \leq i < n$. If the ith pulse is a silence (a dot), then search backwards to find the nearest onset; go back to the end of the rhythm in the event the rhythm is cyclic [57]. Let the nearest onset be at pulse j. Let s be the value of $w_i - w_j$. When s is greater than zero, Longuet-Higgins and Lee say that a syncopation has occurred [106]. The sum of all values s is defined to be the *Longuet-Higgins and Lee Complexity* (or LHL) measure.

We must note one important caveat in the algorithm just described. Consider that a syncopation has just occurred between a silence at pulse i and and onset at pulse j; i.e., $w_i - w_j > 0$. Here, a check must be performed where we skip the silences after pulse i



(a) $2 \cdot 2 \cdot 3$ (b) $2 \cdot 3 \cdot 2$

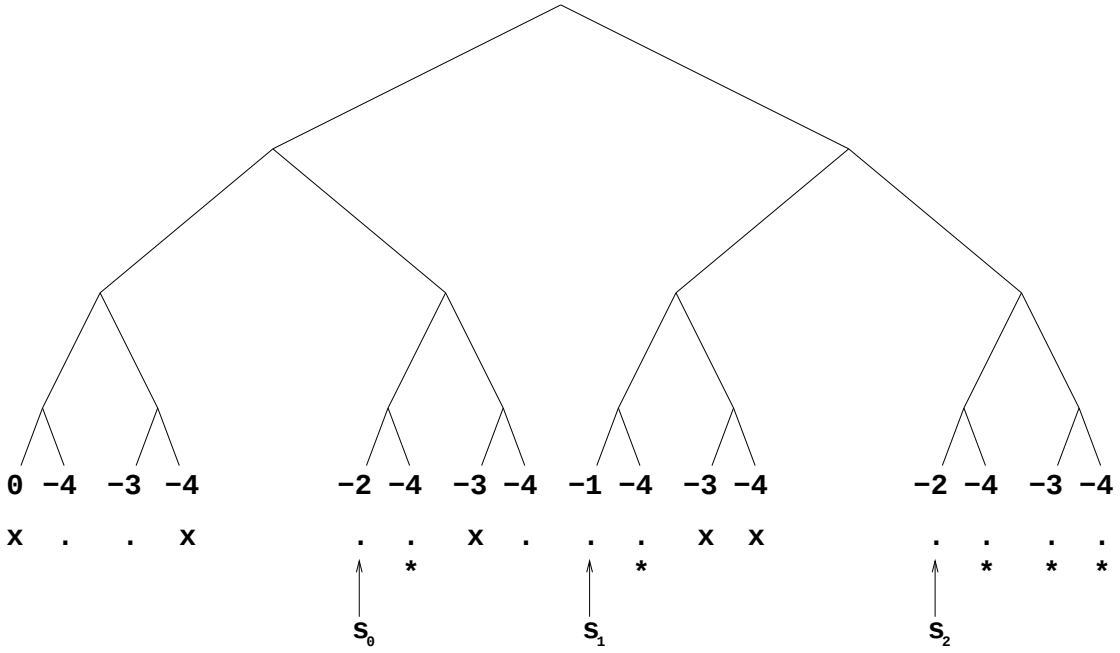
(c) $3 \cdot 2 \cdot 2$

Figure 3.6: *Longuet-Higgins and Lee weighted metrical hierarchies for a rhythm with 12 pulses.*

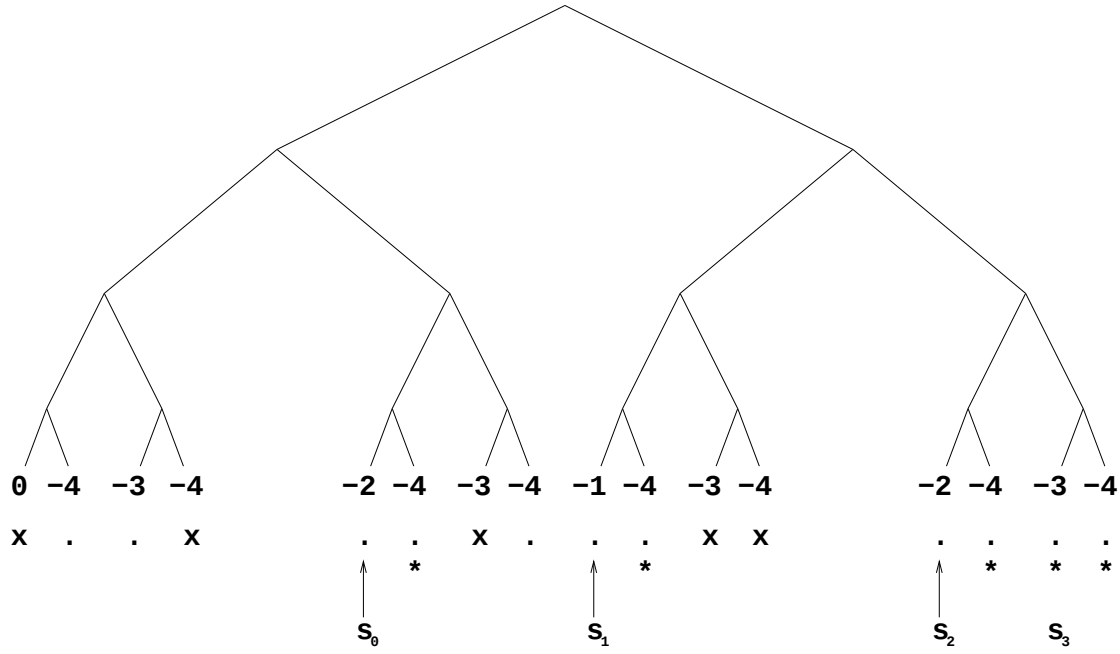
until the weight of a following silence is greater than w_i. This implies that the check stops if we encounter an onset. Once the check terminates, we continue with the algorithm as defined. The reason for this, is the following rule stated by Longuet-Higgins and Lee, "any divided metrical unit consisting entirely of rests, or of notes that are tied together, is replaced by an undivided unit composed of a single rest, or of a single note that is sounded or tied according to whether the first note of the replaced unit was sounded or tied" [106]. The phrase *tied together* indicates that two notes (or rests) are read as a single note (or rest) with the duration being the sum of the two. Thus, the rule implies that silences which follow a silence with a higher metrical weight are considered to be an undivided metrical unit. However, this only applies when there is a syncopation. When there is not a syncopation, we need to check the silences at all metrical levels.

Aside from our implementation of the LHL measure, we consider two others: one by Fitch and Rosenfeld [57], and another by Smith and Honing [150]. Fitch and Rosenfeld's version uses the method described for measuring the complexity; however, they do *not* handle the caveat just described. To see the difference between our implementation and theirs, consider the comparison in Figure 3.7.

In Figure 3.7, we can see that the implementation by Fitch and Rosenfeld in Figure 3.7(b) identifies an additional syncopation marked by s_3, which is really part of the metrical unit of s_2. In the figures, the arrow points to the silence which yields a syncopation with the preceding onset, and the asterisks (*) indicate those silences which are to be



(a) Our implementation.



(b) Fitch and Rosenfeld's.

Figure 3.7: *Comparison of our implementation of the LHL complexity measure in (a), versus Fitch and Rosenfeld's implementation of the LHL complexity measure in (b). The rhythm used here is the* soukous, [x..x..x...xx....].

skipped since they are part of the metrical unit of the silence where the syncopation occurred. In Fitch and Rosenfeld's version we see that the silences with an asterisk are not skipped and thus they add an additional syncopation at s_3. This yields the complexity of the *soukous* to be $s_0 + s_1 + s_2 + s_3 = 2 + 2 + 2 + 1 = 7$. On the other hand, our version shown in Figure 3.7(a), respects the metrical unit and yields, $s_0 + s_1 + s_2 = 2 + 2 + 2 = 6$.

Regarding Smith and Honing's [150] version, their code is not available; however, from a personal communication [151] the results of their algorithm were obtained. From this, their measure is assumed to operate in the same manner as the one we describe here; however, the only difference is that they define a syncopation to occur if $w_i - w_j + 1 > 1$. Thus in the example with the *soukous*, the same syncopations are detected, but the values are $s_0 + s_1 + s_2 = 3 + 3 + 3 = 9$. Our version of the LHL Complexity measure is abbreviated as *lhl*, Fitch and Rosenfeld's implementation is abbreviated *fitch*, and Smith and Honing's implementation is abbreviated *smith*.

3.2 Pattern Matching

The next collection of measures is based on patterns within a musical rhythm which determine its complexity. Here we present a measure introduced by Pressing [125, 126], also a measure developed in the work of Tanguiane [157, 158], and finally a complexity measure presented by Keith [86].

3.2.1 Pressing's Cognitive Complexity

Similar to the idea of a metrical hierarchy, Pressing describes a measure which divides a rhythm in a hierarchical way, and then searches each division for *syncopated* patterns, which determine the rhythm's complexity [125, 126].

First, consider how a rhythm may be divided in a hierarchical manner. The method suggested by Pressing's work [125, 126], is that a rhythm is split into its simplist divisions; i.e., again we shall divide the rhythm based on the metrical hierarchy as seen in Toussaint's Metrical Complexity and the Longuet-Higgins and Lee measure. Thus, we find the prime factorization of the number of pulses to yield our hierarchical division. For example, when the number of pulses $n = 16$, our prime factorization is: $2 \cdot 2 \cdot 2 \cdot 2$, and so we first divide the 16 pulses by 2, which yields two sub-patterns of 8 pulses. Then we divide those two sub-patterns by 2 again, which yields four sub-patterns of 4 pulses, and so on, until we use up the prime factors. Consider Figure 3.8, which shows how the *clave son* rhythm is divided into hierarchical levels.

Once the rhythm is divided as seen in Figure 3.8, Pressing defines 6-types of patterns which he weights according to how syncopated they are [125]. Pressing describes that such patterns have a cognitive cost, and so we say that this is *Pressing's Cognitive Complexity* measure. In each *sub-rhythm*, we look for such patterns defined by Pressing. By sub-rhythm

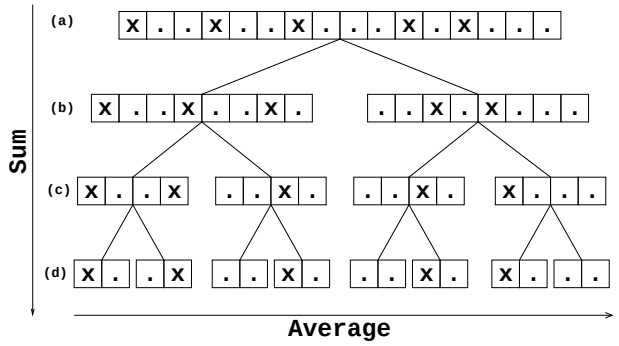


Figure 3.8: *Example of metrical divisions in Pressing's complexity measure on the* clave son, [x..x..x...x.x...].

we mean one of the chunks of the rhythm at a metrical level. For example, in metrical level labeled by (d) in Figure 3.8, we see there are 8 chunks (i.e., 8 sub-rhythms) scanning left to right. We consider each sub-rhythm to see if one of the following patterns is a match.

a. *Filled*: a pattern where each pulse of the sub-rhythm has an onset, or sub-sub-rhythms below the current one all begin with an onset.
b. *Run*: a pattern where an onset starts the sub-rhythm and is followed by a sequence of onsets.
c. *Upbeat*: a pattern where the last pulse in a sub-rhythm contains an onset and the first pulse of the next sub-rhythm, to the right, also contains an onset.
d. *Subbeat*: this is stated, but not defined by Pressing [125, 126], and thus, we exclude it here.
e. *Syncopated*: a pattern where an onset begins and ends off the beat, which means that a sub-sub-rhythm of the current sub-rhythm does not begin with an onset.
f. *Null*: this is when there are no onsets in the sub-rhythm, or when there is only one onset on the first pulse of the sub-rhythm.

For each syncopation type, a weight is assigned: filled is weight 1, run is weight 2, upbeat is weight 3, and syncopated is weight 5. Also, null is weight 0. Referring again to Figure 3.8, we see that going across each metrical level, the weights are averaged for each sub-rhythm. Then going down the metrical levels, the average weights are then summed. This sum value is Pressing's Cognitive Complexity. Pressing describes in his work [125, 126], only a binary division of the rhythm, and presents no material for ternary divisions. Thus, for this measure to accommodate rhythms with 12 pulses, we follow the same divisions as seen in the Metrical Complexity and LHL measures, and so an average among the unique permutations of the prime factorization is taken. Also, we abbreviate Pressing's measure as *pressing*.

3.2.2 Tanguiane's Complexity

Tanguiane introduces a rhythmic model that is based on low-level and high-level pattern interactions during the perception of a rhythm. Low-level patterns represent the actual data, thus in the case of a rhythm, this would be the onsets with their durations. High-level patterns represent the relationships between the low-level patterns and other properties of rhythm, such as tempo [158]. The goal of the proposed model is to represent a rhythm with the least amount of *overall complexity* [158]. The term overall complexity is used in the sense of memory; i.e., the model which represents the rhythm using the least amount of memory, is the model with least complexity.

Here, we are only concerned with the low-level complexity which represents the amount of redundancy in a rhythm [158]. At the low-level, patterns within a rhythm are reduced in terms of *musical elaboration*. Those patterns which are not an elaboration of another are known as *root patterns*. The number of such root patterns is the low-level complexity. The more root patterns there are, the more complex the rhythm [157].

Let us refer to this low-level complexity as *Tanguiane's Complexity* measure, and define some of the terms introduced in the previous paragraph. Thus, we consider musical elaborations and root patterns. An elaboration is a subdivision of a rhythmic pattern [158], which Tanguiane derives from Mont-Reynaud's and Goldstein's notion of an unconstrained elaboration [110]. The pattern which is used for elaboration is called the root pattern [157]. Consider the following method to determine whether one pattern is an elaboration of another, and which is the root pattern.

Since the rhythms Tanguiane describes are taken to be in the binary notation, where a 1 marks an onset and a 0 marks a silence, one way to check for root patterns is by using the bitwise and operation. Here we denote this operation by $\wedge$. The patterns are assumed to both have the same length and start with an onset (a 1), because if there is no duration for the pattern, there is nothing to elaborate [146]. Also, recall that bitwise and compares each pattern, bit-by-bit and yields a 1 when both bits are one, 0 otherwise. Consider an example where a binary strings $r = 1000$ and $s = 1010$.

$$1000 \wedge 1010 = 1000 \tag{3.3}$$

If $r \wedge s$ is equal to r then, without loss of generality, s is considered to be the elaboration of r; moreover r is the root pattern. We see this is the case in Equation (3.3). To determine Tanguiane's Complexity, the operation described is used between all pairs of patterns to count the number of root patterns at a given metrical level of the rhythm. The maximum number of root patterns over each metrical division level is Tanguiane's Complexity [157].

An implementation of this measure is presented by Shmulevich and Povel [146]. How-

ever, as they point out, "[Tanguiane's] definition of complexity is problematic since some structural [metrical] levels may not lend themselves to subdivision" [146]. Thus, they constrain the metrical levels by only considering those which subdivide the pattern such that each division begins with an onset, in accordance with Tanguiane. As an example, consider Figure 3.9. Here we see that only metrical levels (b) and (c) may be used. Level (d) may not be used because there are sub-rhythms which begin with a 0.

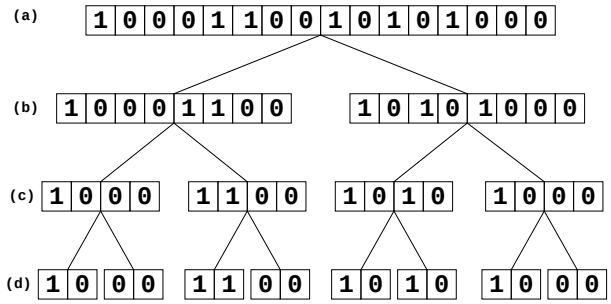


Figure 3.9: *Metrical levels and sub-patterns of an imaginary rhythm indicating that Tanguiane's Complexity measure cannot be used for each level.*

Shmulevich and Povel term these levels, such as (b) and (c), *allowable*. In addition, their method provides the option of further restricting the metrical levels by using a subset of allowable levels determined a priori [146]. We follow this implementation abbreviating it as *tmmax* when the allowable level with the maximum complexity is taken, and *tmavg* when the average of each allowable level taken.

Though, to generalize Tanguiane's Complexity measure, we also introduce a variant which allows sub-rhythms to start with a 0, thus the constraint that an allowable level admits sub-rhythms beginning with a 1, is dropped. In this variant we separate those sub-patterns which start with a 1 from those which start with a zero, considering all possible metrical levels. Once separated, Tanguiane's Complexity is calculated as normal with those sub-patters which start with a 1. However, for those which do not start with a 1, a preliminary step is taken: *complementation*.

The complement is not new for rhythms (binary patterns) [129, 159, 172]. Thus, we take the complement of a rhythm to be a new rhythm where all the 1s are replaced by 0s and all the 0s are replaced by 1s. Essentially, we perform a bitwise not operation; e.g., the complement of $[0100]$ is $[1011]$.

Using the complement operation, we find the complement of all sub-rhythms, which do not start with a 1, and then apply Tanguiane's Complexity as usual to the complemented sub-patterns. Thus, we arrive at a value for Tanguiane's Complexity for the sub-patterns beginning with a 1 and also (with the complement operation) those sub-patterns beginning with a 0. We then add the two values to determine Tanguiane's Complexity for the metrical

level. We call this variant *Tanguiane's Unconstrained Complexity*. Also, when the maximum value from the metrical levels is taken, we abbreviate this as *tmumax*, and when an average is taken, we abbreviate this as *tmuavg*. Also, note that to handle rhythms which will yield more than one metrical structure, such as those rhythm with 12 pulses, the average between the structures is taken for the complexity.

3.2.3 Keith's Complexity

Keith proposes a measure for syncopation based on three rhythmic events: hesitation, anticipation, and syncopation. A hesitation occurs when an onset begins on a beat and ends off a beat, an anticipation occurs when an onset begins off a beat and ends on a beat, and a syncopation occurs when an onset begins off a beat and ends off a beat [67, 86]. The rhythmic event patterns are then used to compute the syncopation of a rhythm. Consider the following five step procedure of Keith's syncopation measure.

First, let $\mathbf{r}$ be a rhythm with n pulses, k onsets, and index each pulse of $\mathbf{r}$ from 0 to $n-1$. Second, pick any onset and let i be the pulse where that onset is sounded, let j be the pulse of the onset following the one which was picked. Compute the duration $\delta = j - i$ and define $\hat{\delta}$ to be δ rounded down to the nearest power of 2. The value $n/\hat{\delta}$ gives the number of evenly spaced beats among the n pulses, i.e., the meter. Third, if i modulo $\hat{\delta}$ is congruent to 0 then i is on a beat, otherwise i is off a beat. If j modulo $\hat{\delta}$ congruent to 0 then j is on a beat, otherwise j is off a beat. Fourth compute the value of syncopation s as in Equation (3.4), weighting a hesitation as 1, anticipation as 2, and syncopation as 3. Fifth, repeat at step two, picking an onset not already chosen, and once all onsets have been picked, sum the weights s from step four to obtain *Keith's Complexity* measure. Note that Keith's Complexity measure is abbreviated *keith*.

$$s = \begin{cases} 0 & \text{if } i \text{ is on beat and } j \text{ is on beat} \\ 1 & \text{if } i \text{ is on beat and } j \text{ is off beat} \\ 2 & \text{if } i \text{ is off beat and } j \text{ is on beat} \\ 3 & \text{if } i \text{ is off beat and } j \text{ is off beat} \end{cases} \tag{3.4}$$

As an example, consider $\mathbf{r}$ to be the clave son $[\mathsf{x}\,.\,.\,\mathsf{x}\,.\,.\,\mathsf{x}\,.\,.\,.\,\mathsf{x}\,.\,\mathsf{x}\,.\,.\,.]$ where $n = 16$ and $k = 5$. Let $i = 0$ and $j = 3$, which represents the beginning and end of the first onset. We have $\delta = j - i = 3$ and so $\hat{\delta} = 2$, since 2^1 is the closest power of two less than 3. This tells us that there are $n/\hat{\delta}$, or 8, beats evenly spaced in the 16 pulses. Thus, beginning with pulse 0, we have a beat every other pulse: $0, 2, 4, \ldots$. If we compute i modulo $\hat{\delta}$ to be congruent to 0, then i begins on a beat. However, if computing j modulo $\hat{\delta}$ is congruent 1, then j is off a beat; i.e., i does not end on beat. Using Equation (3.4), we have $s = 1$, which is interpreted to mean a hesitation. This process is then repeated for each onset. Note the

end of the last onset is taken to be where the first onset begins since our rhythm is cyclic. Thus, when $i > j$, we take $j + n$ for ease of computation. After computing the measure for each onset, the total syncopation for $\mathbf{r}$ is 6. We had events: hesitation, anticipation, and syncopation.

3.3 Distance

Here, the mathematical notion of distance is used to determine how complex a rhythm is. Essentially, the following two measures compare rhythms to meters with even beat distributions, and then use a measurement of distance to judge complexity. First we present a measure introduced by Toussaint et al. [28, 40, 168, 169], and then one by Gómez et al. [67].

3.3.1 Directed Swap Distance

The *Directed Swap Distance* has previously been shown to be useful for measuring the similarity between rhythmic patterns [40, 168, 169]. Moreover, based on a study of rhythm similarity measures [173], a reduced version of the Directed Swap Distance, the *Swap Distance*, performed the best along with the chronotonic distance [40, 173]. Here, the Directed Swap Distance is used instead of the swap distance because the rhythms being compared may have a different number of onsets; the swap distance relies on rhythms with the same number of onsets. In terms of complexity, one rhythm is said to be more complex than another if the Directed Swap Distance of the first rhythm is greater than the second when transforming both rhythms to a simple pattern, such as a meter. Thus, we have a complexity measure based on the dissimilarity a rhythm compared to a fixed meter.

We will touch on complexity below, but first let us gain a flavor for the algorithm of the Directed Swap Distance by considering the Swap Distance. In general, a *swap* may be thought of as an interchange of position between elements. This notion is present in sorting algorithms [37], as the *shift* operation in the *Fuzzy Hamming Distance* [10], and as a one dimensional version of the *Earth Mover's Distance*, where all points have an equal weight [25, 176]. Thus, the term swap has a variety of interpretations.

Let us use the interpretation that a swap is an interchange between two *adjacent* elements. The Swap Distance is the minimum number of swaps required to transform one pattern into another. More formally, consider the binary alphabet used to represent a musical rhythm, $A = \{\mathsf{x}, .\}$, where A^* is the set of all binary strings over the alphabet. Let $\mathbf{r}, \mathbf{s} \in A^*$ be strings (in box notation), which represent the *clave son* and *bossa nova* rhythms, respectively.

$$\mathbf{r} = [\mathsf{x}..\mathsf{x}..\mathsf{x}...\mathsf{x}.\mathsf{x}...] \tag{3.5}$$

$$\mathbf{s} = [\mathsf{x}..\mathsf{x}..\mathsf{x}...\mathsf{x}..\mathsf{x}..] \tag{3.6}$$

Let each pulse be indexed 0 through 15, thus r_0 is an 'x'. If we swap r_{12}, an 'x', with r_{13}, a '.', then we arrive at pattern $\mathbf{s}$. Since only one swap was required, the Swap Distance is equal to 1. If we consider the index position of each onset (each 'x') in $\mathbf{r}$ and $\mathbf{s}$, then we can create two vectors $\mathbf{p} = (0, 3, 6, 10, 12)$ and $\mathbf{q} = (0, 3, 6, 10, 13)$ and define the Swap Distance calculation as follows where k is the length of $\mathbf{p}$ and $\mathbf{q}$.

$$d_{swap} = \sum_{i=0}^{k-1} |p_i - q_i| \tag{3.7}$$

The absolute difference between each onset position will yield the number of swaps it takes to align all the onsets of one rhythm to the other. Now, we must consider when the number of onsets k is different between two rhythms. Let $\mathbf{u}, \mathbf{v} \in A^*$ be rhythms where $\mathbf{u}$ has more onsets than $\mathbf{v}$. In this case, finding the minimum number of swaps required to align each onset in $\mathbf{u}$ to some onset in $\mathbf{v}$, where each onset in $\mathbf{v}$ must have at least one onset from $\mathbf{u}$ aligned to it. This is known as the Directed Swap Distance [28]. This measure was originally proposed by Toussaint, and applied to rhythm similarity by Báñez, et al. [40].

Colannino and Toussaint [28] propose an algorithm to compute the Directed Swap Distance which has a worst case running time of $O(n^2)$. The main idea of the algorithm is to change the problem into finding the shortest path in a single source acyclic weighted graph. The vertices represent possible assignment choices between the binary patterns, and the edges are weighted by the cost of assigning the onsets of one pattern to the other [28]. Once the graph is constructed, the shortest path may be found using Dijkstra's algorithm [30], for example. Additionally, Colannino et al. [26] reduced the time complexity of the computing the Directed Swap Distance to $O(n)$ for binary patterns, and $O(n\ log\ n)$ for real numbers. The correctness of such algorithms have been shown elsewhere [26, 28], but consider the following which exemplifies how the Directed Swap Distance (d_{dswap}) can be used as a measure of rhythmic complexity.

Let $\mathbf{r}, \mathbf{s}, \mathbf{t} \in A^*$ be strings in box notation, all with length $n = 16$. Take $\mathbf{r}$ to represent the fixed meter where we have a beat at every fourth pulse, starting at pulse 0; i.e., at positions 0,4,8, and 12. Also, take $\mathbf{s}$ to represent the *shiko* rhythm and $\mathbf{t}$ to represent the *gahu* rhythm. These rhythms were chosen because the *gahu* rhythm is considered to be more complex than the *shiko* rhythm [166].

$$\mathbf{r} = [\mathsf{x}\ldots\mathsf{x}\ldots\mathsf{x}\ldots\mathsf{x}\ldots] \tag{3.8}$$

$$\mathbf{s} = [\mathsf{x}\ldots\mathsf{x}.\mathsf{x}\ldots\mathsf{x}.\mathsf{x}\ldots] \tag{3.9}$$

$$\mathbf{t} = [\mathsf{x}..\mathsf{x}..\mathsf{x}\ldots\mathsf{x}\ldots\mathsf{x}.] \tag{3.10}$$

Let $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ be integer vectors which indicate the onset positions in $\mathbf{r}, \mathbf{s}$, and $\mathbf{t}$.

$$\mathbf{u} = (0, 4, 8, 12) \tag{3.11}$$

$$\mathbf{v} = (0, 4, 6, 10, 12) \tag{3.12}$$

$$\mathbf{w} = (0, 3, 6, 10, 14) \tag{3.13}$$

By calculating the directed swap distance between $\mathbf{v}$ and $\mathbf{u}$ and then between $\mathbf{w}$ and $\mathbf{u}$, we arrive at $d_{dswap}(\mathbf{v}, \mathbf{u}) = 4$ and $d_{dswap}(\mathbf{w}, \mathbf{u}) = 7$. Thus, we see that the *gahu* rhythm represented by the onset positions of $\mathbf{w}$ yields a larger Directed Swap Distance to a fixed meter than the *shiko* rhythm which is represented by the onset positions of $\mathbf{v}$. Therefore, the *gahu* rhythm is more complex than the *shiko* rhythm by the Directed Swap Distance, which agrees with previous analysis [166]. Thus, we implement the Directed Swap Distance using fixed meters of 2 beats, 4 beats, and 8 beats for rhythms with 16 pulses, and fixed meters of 2 beats, 3 beats, 4 beats, and 6 beats for rhythms with 12 pulses. In addition, we define a version of the measure which takes the average over all corresponding fixed meters for a given number of pulses. The abbreviation is *dswap*, for the average, and *dswap#*, where $\#$ is the number of pulses between each beat, so *dswap2* means there are 8 beats.

3.3.2 Weighted Note-to-Beat Distance

The *Weighted Note-to-Beat Distance* [67, 68] (or more simply WNBD) is a distance measure for syncopation. The WNBD measure computes the distance between onsets and nearest beats in a rhythm, and then assigns a weighted score based on that distance. The sum of all such scores divided by the number of onsets is the measure of syncopation [67]. More formally, let $\mathbf{r}$ be a rhythm, with n pulses, k onsets, and m be the number of beats in the meter. Let $\mathbf{e}$ be a vector of length m where each beat $e \in \mathbf{e}$ is evenly spaced among the pulses of $\mathbf{r}$, starting at pulse 0 in the interval $[0, n)$, where $[,)$ is defined to be a left-inclusive, right-exclusive interval.

Let x be an onset in $\mathbf{r}$, then pick e_i, e_{i+1}, e_{i+2} to be beats. For cyclic patterns, we take i (modulo m) [67]. Consider three steps to compute WNBD.

First, calculate the distance of (x, e_i) and (x, e_{i+1}). Pick the smaller distance. Let this be $T(x)$. The distance is measured as the number of pulses between the onset and the nearest beat, which is then divided by m. Second, assign weight $D(x)$ as follows.

$$D(x) = \begin{cases} \frac{1}{T(x)} & \text{if } x \text{ ends before or on } e_{i+1} \\ \frac{2}{T(x)} & \text{if } x \text{ ends after } e_{i+1} \text{ but before or on } e_{i+2} \\ \frac{1}{T(x)} & \text{if } x \text{ ends after } e_{i+2} \\ 0 & \text{if } T(x) = 0 \end{cases} \tag{3.14}$$

Third, sum the weight $D(x)$ for each onset x and then divided by the number of onsets k. This value is the measure of syncopation for rhythm $\mathbf{r}$ [67]. Regarding the weights, the highest value is given when the onset x ends after e_{i+1}, but before or on e_{i+2}. The reason for this is because a beat (namely e_{i+1}) is unsounded, which is theorized to produce a strong syncopation. For the other two weight values, more weight is given to onsets that are a small distance away from a beat. When an onset lands close-to, or on, the mid-point between two beats, the lesser weight is given (greater than 0) [67].

Consider an example of the WNBD measure given the cyclic *clave son* rhythm in box notation, [x..x..x...x.x...], with $n = 16$, $k = 5$, and $m = 4$. Hence, we have four beats, (e_1, e_2, e_3, e_4) at pulses $(0, 4, 8, 12)$. The first onset x_0 is on the first beat, e_0, and so $D(x_0) = 0$. The second onset is distance 3 from e_0 and distance 1 from e_1, so $T(x_1) = 1/4$. We have that x_1 ends after e_1, but before e_2. Therefore, according to Equation 3.14, $D(x_1) = 2/T(x_1)$. Onset x_2 presents the same case, and so $T(x_2) = 1/2$ and $D(x_2) = 2/T(x_2)$. The onset x_3 is equidistant from beats e_3 and e_4, which yields $T(x_3) = 1/2$. Also, x_3 ends on e_3 which yields $D(x_3) = 1/T(x_3)$. Finally, the onset x_4 gives $D(x_4) = 0$, since the onset is sounded on strong beat e_3. Taking the sum and dividing by k, we arrive at the following value shown in Equation 3.15.

$$WNBD(R) = \frac{1}{k} \sum_{i=0}^{m-1} D(x_i) = 14/5 \tag{3.15}$$

Thus we implement the WNBD measure, and because the number of beats is a chosen parameter, we use all appropriate values for meters of rhythms with 16 and 12 pulses. Thus we have 2 beats, 4 beats, 8 beats, and an average of all three for rhythms with 16 pulses, and we have 2 beats, 3 beats, 4 beats, 6 beats, and an average for rhythms with 12 pulses. The abbreviations are $wnbd$ for the average and $wnbd\#$, where $\#$ denotes the number of pulses between each beat, so *wnbd6* means there are 2 beats.

3.4 Information Entropy

The complexity measures presented in this section rely on calculating the information entropy, in the sense of Shannon's entropy [64, 143]. Below, Equation (3.16) shows the formula for calculating the information entropy of a discrete random variable. Equation (3.17) shows the formula for calculating the joint entropy between two discrete random variables [31], and Equation (3.18), shows the formula for calculating the conditional entropy of two discrete random variables.

The conventions used here are that we assume $\mathcal{X}$ and $\mathcal{Y}$ are discrete random variables which take on values $x \in X$ and $y \in Y$, respectively. Moreover, $p(x)$ is written for $\Pr\{\mathcal{X} = x\}$ (i.e., the probability that $\mathcal{X}$ takes on x), $p(x, y)$ is written for $\Pr\{\mathcal{X} = x \text{ and } \mathcal{Y} = y\}$

(i.e., the joint probability that $\mathcal{X}$ takes on y and $\mathcal{Y}$ takes on y), and $p(y|x)$ is written for $\Pr\{\mathcal{Y} = y \text{ given } \mathcal{X} = x\}$ (i.e., the conditional probability that $\mathcal{Y}$ takes on y, given that $\mathcal{X}$ has taken on x). Consider the equations for information entropy, joint entropy, conditional entropy, and then the following sections presenting three measures based on the notion of information from Vitz [179] and Vitz and Todd [180]. Note that we define $0\ log_2\ 0$ and $1\ log_2\ 1$ to be 0.

$$H(\mathcal{X}) = -\sum_{x \in X} p(x)\ log_2\ p(x) \tag{3.16}$$

$$H(\mathcal{X}, \mathcal{Y}) = -\sum_{x \in X} \sum_{y \in Y} p(x)\ p(x, y)\ log_2\ p(x, y) \tag{3.17}$$

$$H(\mathcal{X}|\mathcal{Y}) = -\sum_{x \in X} \sum_{y \in Y} p(y, x)\ log_2\ p(y|x) \tag{3.18}$$

3.4.1 H(k-span) Complexity

The *H(k-span) Complexity* presented by Vitz [179] in 1968 models the perception of stimuli in the form of binary patterns, providing a measure of complexity. This model assumes that a subject evaluates the uncertainty of the transition between sub-patterns of the binary string [179]. The $H(k$-span) model evaluates uncertainty starting with H_0 (order zero) and then progresses to evaluate each sequential order: H_1, H_2, ..., H_m, where $H_m = 0$ [179]. Thus, we can say that $H(k$-span) pattern complexity is the total of the average uncertainties evaluated at each order up to H_m. Consider Equation 3.19.

$$H(k\text{-span}) = \sum_{k=0}^{m} H_k \tag{3.19}$$

At each order, H_k, the uncertainty is evaluated in the following manner. Let $\mathbf{r}$ be a cyclic (infinitely repeating) binary string of length n. First, find all 2^k binary strings of length k for $1 \leq k \leq m$; e.g., when $k = 2$, then we have binary strings $00, 01, 10$, and 11. Let the set S contain these strings. Second, for each binary string $s \in S$, count the number of times s appears as a substring of $\mathbf{r}$, let this be u. Let v be the number of substrings length k, in $\mathbf{r}$. Thus, the probability of obtaining substring s is u/v. Third, since we are using a binary alphabets of zeros and ones, we find the probability that the bit following substring s is a zero or a one. So, for each instance of s found in $\mathbf{r}$, look at the next bit. Count the number of times the next bit is a one, let this be z. Count the number of times the next bit is zero, let this be w. Now, we have the probability that the bit following s is a one, being z/u, with probability that the next bit is a zero, being w/u.

Fourth, we may use these probabilities to calculate the information entropy over each substring s and each bit 1 or 0 that may follow s. Thus, the standard formulas for entropy and joint entropy are used. Review Equations (3.16) and (3.17), respectively, which have previously been derived from [8, 31, 143].

Equation (3.17) can be used to measure the entropy. In the equation, $\mathcal{X}$ is a discrete random variable. Here we say that $\mathcal{X}$ may take on values which are binary strings, such as s, also $\mathcal{Y}$ is a discrete random variable which may take on a one or a zero. Therefore in our example, $p(x)$ when $\mathcal{X}$ takes on s is $p(s) = u/v$, $p(x, y)$ when $\mathcal{X}$ takes on s and $\mathcal{Y}$ takes on 1 is $p(s, 1) = z/u$, and $p(x, y)$ when $\mathcal{X}$ takes on s and $\mathcal{Y}$ takes on 0 is $p(s, 0) = w/u$.

However, instead of this tedious calculation of entropy and joint entropy, it has been observed that the $H(k$-span) is equivalent to log_2 of the length of the binary string, as stated by Vitz [179]. Thus this thesis implements this simpler computation for the $H(k$-span) Complexity, which is abbreviated as *hk*.

3.4.2 H(run-span) Complexity

The *H(run-span) Complexity* measure is similar to the $H(k$-span), except a preliminary step is taken which codes adjacent like elements into *runs*. A run is any event repeated some number of times and preceded and followed by an event other than itself [179]. For example, in the pattern 0001, the sequence 000 is a run. Once the preliminary step of coding runs is complete, the H(run-span) is exactly the same as the $H(k$-span) measure, but here the alphabet grows from a binary alphabet to one containing an element representing each run in the binary string; e.g., 000 would be considered an element in the alphabet, as would 1, but not 0.

Again the H(run-span) may be calculated using the formulas for entropy and joint entropy as shown in Equations (3.16) and (3.17), since it retains the same definition as $H(k$-span), but again there is a simple computation for the H(run-span) complexity. Vitz points out that the complexity may be found by taking log_2 of the number of runs found in the pattern [179]. Thus, we implement this measure using the computation on the number of runs. Note that this measure is abbreviated as *hrun*.

3.4.3 Coded Element Processing System

Vitz and Todd [180] present a procedure that can be used to calculate the complexity of cyclic binary patterns presented as a sequence of tones and non-tones [180], or in our case onsets and silences. Their model specifies axioms which describe the perceptual encoding process. This model is called the *Coded Element Processing System* or CEPS [180]. CEPS is a parameter free model which makes the assumption that subjects closely approximate this model when dealing with patterns of simple to moderate complexity [180].

CEPS examines each pattern in a hierarchical manner pertaining to levels of perceptual

organization. The pattern is coded into stimulus elements and response elements, where the response elements in-turn become the stimulus elements [180]. The processes of coding the pattern from single elements to larger elements continues until the entire pattern is constructed [180]. Consider the axioms below which summarize the model [180].

Axiom 3.4.1. *The stimuli (1s or 0s) are represented as the Code Level* 1 *elements, and denoted as* e_1.

Axiom 3.4.2. *The* e_1 *elements are coded into runs (adjacent 1s or 0s), which represent Code Level* 2 *elements, denoted as* e_1^n *where* n *is the number of elements in the run.*

Axiom 3.4.3. *Adjacent* e_1^n *elements are grouped together to form a new element called a* composite*, denoted as* e_2*. Elements are grouped into a composite is until an* e_1 *element is encountered that already exists in the composite. At this point, a new composite is started. Note that in the case of cyclic patterns, there may be different configuration of composites depending upon where one starts. To handle this, the configuration with the smallest number of composites is chosen; otherwise, the system does not distinguish between configurations [180].*

Axiom 3.4.4. *Group the* e_2 *elements into runs called Code Level 4 elements, denoted as* e_2^n.

Axiom 3.4.5. *Apply operations of Axioms 3.4.3 and 3.4.4 to form new composite elements and new runs.*

Axiom 3.4.6. *The stopping criterion is when the pattern itself is formed as a coded element.*

The axioms determine an algorithm for building a hierarchy of elements grouped into larger structures, which depend on the characteristics of the pattern. Vitz and Todd state that characteristics such as runs of like events, alternations, and runs of alternations, are commonly coded by humans [180].

To calculate the complexity of the sequence, Vitz and Todd measure two uncertainty values at each Code Level in the algorithm. Uncertainty in this case means *information entropy* [31, 143]. The uncertainty values are the *maximum uncertainty*, H_{max}, and the *joint uncertainty*, H_{joint} [64, 180]. H_{max} measures the uncertainty of the occurrence of random variables, with the assumption that each random variable is uncorrelated. H_{joint} measures the uncertainty of the occurrence of random variables given their actual correlation [64]. The following Equations (3.20) and (3.21), show H_{max} and H_{joint}. Let $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, and $\mathcal{W}$ be discrete random variables. Recall that $H(\cdot)$ and $H(\cdot|\cdot)$ represent standard equations of entropy and conditional entropy [31], as previously presented by Equations (3.16) and (3.18).

$$H_{max}(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{W}) = H(\mathcal{X}) + H(\mathcal{Y}) + H(\mathcal{Z}) + H(\mathcal{W}) \tag{3.20}$$

$$H_{joint}(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{W}) = H(\mathcal{X}) + H(\mathcal{Y}|\mathcal{X}) + H(\mathcal{Z}|\mathcal{X}, \mathcal{Y}) + H(\mathcal{W}|\mathcal{X}, \mathcal{Y}, \mathcal{Z}) \tag{3.21}$$

Let the value H^k be the sum of H^k_{max} and H^k_{joint} at Code Level k. The sum of H at each Code Level represents the complexity of the pattern. This value is termed H_{code}. Equation 3.22 shows this computation where there are n Code Levels [180].

$$H_{code} = \sum_{k=1}^{n} H^k \tag{3.22}$$

To gain a flavor for the CEPS algorithm, consider the following example of the binary pattern $[01101101]$. Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be discrete random variables taking on elements at each Code Level. Let $\mathcal{Y}_1$ and $\mathcal{Y}_2$ be discrete random variables taking on integers greater than or equal to zero. The random variable $\mathcal{X}_1$ takes on an element at the given Code Level, and $\mathcal{X}_2$ takes on the element that follows $\mathcal{X}_1$. Hence, there is an implied dependence between each variable. The random discrete variable $\mathcal{Y}_1$ takes on value for the length of element $\mathcal{X}_1$ takes on, and $\mathcal{Y}_2$ takes on a value for the length of the element $\mathcal{X}_2$ takes on. Hence there is dependence between each $\mathcal{X}_1$ and $\mathcal{Y}_1$ and $\mathcal{X}_2$ and $\mathcal{Y}_2$. Table 3.1 provides what each term means in the Equations (3.20) and (3.21).

Table 3.1: *Meaning of each uncertainty value in Equations* (3.20) *and* (3.21).

Uncertainty Value	Meaning
$H(\mathcal{X}_1)$	Uncertainty that $\mathcal{X}_1$ takes on an element of the pattern
$H(\mathcal{X}_2)$	Uncertainty that $\mathcal{X}_2$ takes on an element of the pattern
$H(\mathcal{Y}_1)$	Uncertainty that $\mathcal{Y}_1$ takes on a length where $\mathcal{X}_1$ is unknown
$H(\mathcal{Y}_2)$	Uncertainty that $\mathcal{Y}_2$ takes on a length where $\mathcal{X}_2$ is unknown
$H(\mathcal{Y}_1\|\mathcal{X}_1)$	Uncertainty that $\mathcal{Y}_1$ takes on a length where the element $\mathcal{X}_1$ is known
$H(\mathcal{X}_2\|\mathcal{X}_1, \mathcal{Y}_1)$	Uncertainty that $\mathcal{X}_2$ transitions to an element where the previous element $\mathcal{X}_1$ is known with a known length $\mathcal{Y}_1$
$H(\mathcal{Y}_2\|\mathcal{X}_1, \mathcal{Y}_1, \mathcal{X}_2)$	Uncertainty of the length $\mathcal{Y}_2$ of the known transition to element $\mathcal{X}_2$ from $\mathcal{X}_1$ with length $\mathcal{Y}_1$

Since the H_{joint} requires calculating joint probabilities, consider the tables below. To begin the process, we use Axiom 3.4.1 which takes each symbol in the pattern as an element called Code Level 1. At this first step we want to calculate H^1_{max} and H^1_{joint}. In the patten $[01101101]$ we have $5/8$ probability that $\mathcal{X}_1$ takes on a 1 and $3/8$ that $\mathcal{X}_1$ takes on a 0.

Similarly $\mathcal{X}_2$ has the same distribution as $\mathcal{X}_1$. Since each element is taken individually, the lengths are 1. The probability that $\mathcal{Y}_1$ and $\mathcal{Y}_2$ take on one is 1. Now calculate H^1_{max}.

$$\begin{aligned} H^1_{max}(\mathcal{X}_1, \mathcal{Y}_1, \mathcal{X}_2, \mathcal{Y}_2) &= H(\mathcal{X}_1) + H(\mathcal{Y}_1) + H(\mathcal{X}_2) + H(\mathcal{Y}_2) \\ &= 2 * (5/8 \; log \; 5/8 + 3/8 \; log \; 3/8) + 2 * (1 \; log \; 1 + 0 \; log \; 0) \\ &= 1.91 \end{aligned} \tag{3.23}$$

To find H^1_{joint} we can use the tables below which show the joint distributions for each calculation. The first value $H(\mathcal{X}_1)$ is know from H^1_{max}. To find $H(\mathcal{Y}_1|\mathcal{X}_1)$, Table 3.2 (a). This means that given a sequence element $\mathcal{X}_1$, we want to know the uncertainty of the length of that element. Since we are on the first step, all lengths are one. Once we have these value, we can find $H(\mathcal{X}_2|\mathcal{X}_1, \mathcal{Y}_1)$ (see Table 3.2 (b)), which is the uncertainty of an element following either a 1 or a 0 element in our pattern. Finally, using Table 3.2 (c), we have the uncertainty value $H(\mathcal{Y}_2|\mathcal{X}_1, \mathcal{Y}_1, \mathcal{X}_2)$, which calculates each transition probability from an element $\mathcal{X}_1$ with a known length to an element $\mathcal{X}_2$ with an unknown length. Thus, we have the following.

$$\begin{aligned} H^1_{joint}(\mathcal{X}_1, \mathcal{Y}_1, \mathcal{X}_2, \mathcal{Y}_2) &= H(\mathcal{X}_1) + H(\mathcal{Y}_1|\mathcal{X}_1) + H(\mathcal{X}_2|\mathcal{X}_1, \mathcal{Y}_1) + H(\mathcal{Y}_2|\mathcal{X}_1, \mathcal{Y}_1, \mathcal{X}_2) \\ &= (5/8 \; log \; 5/8 + 3/8 \; log \; 3/8) + 0 \\ &\quad + (5/8 \; (3/5 \; log \; 3/5 + 2/5 \; log \; 2/5)) + 0 \\ &= 1.56 \end{aligned} \tag{3.24}$$

Table 3.2: *Joint probability tables at Code Level 1 for* [01101101].

(a) $H^1(\mathcal{Y}_1|\mathcal{X}_1)$

$p(x_1)$	Y_1	
	0	1
$p(0) = 3/8$	0	1
$p(1) = 5/8$	0	1

(b) $H^1(\mathcal{X}_2|\mathcal{X}_1, \mathcal{Y}_1)$

$p(x_1, y_1)$	X_2	
	0	1
$p(0,1) = 3/8$	0	1
$p(1,1) = 5/8$	3/5	2/5

(c) $H^1(\mathcal{Y}_2|\mathcal{X}_1, \mathcal{Y}_1, \mathcal{X}_2)$

$p(x_1, y_1, x_2)$	Y_2	
	0	1
$p(0,1,1) = 3/8$	0	1
$p(1,1,0) = 3/8$	0	1
$p(1,1,1) = 2/8$	0	1

We follow Axiom 3.4.2 to produce Code level 2 where the pattern contains elements joined to form runs. Thus we have [0 11 0 11 0 1]. Again we can compute H^2_{max} and H^2_{joint}. The process is exactly the same as before, except each element is considered a run,

and now we have elements of different lengths. Thus, we have the following.

$$H^2_{max} = 2.92 \tag{3.25}$$

$$H^2_{joint} = 1.92 \tag{3.26}$$

Table 3.3: *Joint probability at Code Level 2 for* $[0\ 11\ 0\ 11\ 0\ 1]$.

(a) $H^2(\mathcal{Y}_1|\mathcal{X}_1)$

$p(x_1)$	Y_1	
	1	2
$p(0) = 1/2$	1	0
$p(1) = 1/2$	1/3	2/3

(b) $H^2(\mathcal{X}_2|\mathcal{X}_1, \mathcal{Y}_1)$

$p(x_1, y_1)$	X_2	
	0	1
$p(0,1) = 1/2$	0	1
$p(1,1) = 1/6$	1	0
$p(1,2) = 1/3$	1	0

(c) $H^2(\mathcal{Y}_2|\mathcal{X}_1, \mathcal{Y}_1, \mathcal{X}_2)$

$p(x_1, y_1, x_2)$	Y_2	
	1	2
$p(0,1,1) = 1/2$	1/3	2/3
$p(1,1,0) = 1/6$	1	0
$p(1,2,0) = 1/3$	1	0

The next step in the algorithm produces Code level 3, where we form composites, which are essentially a grouping of runs. The algorithm produces two possible composites, $[(0\ 11)(0\ 11)(0\ 1)]$ or $[(11\ 0)(11\ 0)(1\ 0)]$; however, each yields the same probabilities, so let us choose $[(0\ 11)(0\ 11)(0\ 1)]$ for our example. At this step, the elements are the composites (in parentheses), and the length of a composite is the number of elements from Code Level 2, which were used to construct the composite. For $(0\ 11)$ we used two elements from Code Level 2: 0 and 11. Hence, $(0\ 11)$ has length 2. We can calculate H^3_{max}, and using Table 3.4 below, we can find H^3_{joint}.

$$H^3_{max} = 1.84 \tag{3.27}$$

$$H^3_{joint} = 0.66 \tag{3.28}$$

Table 3.4: *Joint probability at Code Level 3 for* $[(0\ 11)(0\ 11)(0\ 1)]$.

(a) $H^3(\mathcal{Y}_1|\mathcal{X}_1)$

$p(x_1)$	Y_1
	2
$p(0\ 11) = 2/3$	1
$p(0\ 1) = 1/3$	1

(b) $H^3(\mathcal{X}_2|\mathcal{X}_1, \mathcal{Y}_1)$

$p(x_1, y_1)$	X_2	
	$(0\ 11)$	$(0\ 1)$
$p(0\ 11, 2) = 2/3$	1/2	1/2
$p(0\ 1, 2) = 1/3$	1	0

(c) $H^3(\mathcal{Y}_2|\mathcal{X}_1, \mathcal{Y}_1, \mathcal{X}_2)$

$p(x_1, y_1, x_2)$	Y_2
	2
$p(0\ 11, 2, 0\ 11) = 1/3$	1
$p(0\ 11, 2, 0\ 1) = 1/3$	1
$p(0\ 1, 2, 0\ 11) = 1/3$	1

Following Axiom 3.4.4 we produce runs of the composites. Thus we have Code Level 4 as the pattern $[\{(0\ 11)(0\ 11)\}\{(0\ 1)\}]$. The curly braces indicate the new runs that were

formed from the composites. Similar to previous steps, we can calculate the H^4_{max} and then use Table 3.5 to find H^4_{joint}.

$$H^4_{max} = 4.00 \tag{3.29}$$

$$H^4_{joint} = 0.00 \tag{3.30}$$

Table 3.5: *Joint probability at Code Level 4 for* $[\{(0\ 11)(0\ 11)\}\{(0\ 1)\}]$.

(a) $H^4(\mathcal{Y}_1|\mathcal{X}_1)$

$p(x_1)$	Y_1	
	1	2
$p(\{(0\ 11)(0\ 11)\}) = 1/2$	0	1
$p(\{(0\ 1)\}) = 1/2$	1	0

(b) $H^4(\mathcal{X}_2|\mathcal{X}_1, \mathcal{Y}_1)$

$p(x_1, y_1)$	X_2	
	$\{(0\ 11)(0\ 11)\}$	$\{(0\ 1)\}$
$p(\{(0\ 11)(0\ 11)\}, 2) = 1/2$	0	1
$p(\{(0\ 1)\}, 1) = 1/2$	1	0

(c) $H^4(\mathcal{Y}_2|\mathcal{X}_1, \mathcal{Y}_1, \mathcal{X}_2)$

$p(x_1, y_1, x_2)$	Y_2	
	1	2
$p(\{(0\ 11)(0\ 11)\}, 2, \{(0\ 1)\}) = 1/2$	1	0
$p(\{(0\ 1)\}, 1, \{(0\ 11)(0\ 11)\}) = 1/2$	0	1

The next step will form the original pattern (by making a composite), thus we apply the stopping criterion of Axiom 3.4.6. We can calculate the H_{code} for $[01101101]$.

$$\begin{aligned} H_{code} &= \sum_{k=1}^{4} H^k = \sum_{k=1}^{4} H^k_{max} + H^k_{joint} \\ &= 1.91 + 1.56 + 2.92 + 1.92 + 1.84 + 0.66 + 4.00 + 0 \\ &= 14.81 \end{aligned} \tag{3.31}$$

Thus we've shown an example calculation for CEPS. This has been implemented in this thesis as a complexity measure for musical rhythms. The abbreviation used is *ceps*.

3.4.4 Lempel-Ziv Coding

A popular method for data compression is the Lempel-Ziv algorithm [98, 186, 187, 31]. The compression of a sequence by the Lempel-Ziv algorithm asymptotically approaches the information entropy of that sequence [31, 187]. Thus, compression may be used to judge the complexity of a sequence. The notion is that a simple sequence will be compressed more easily. The Lempel-Ziv algorithm has also been used to quantify the complexity of a musical rhythm [144, 146, 166]. The main idea of the algorithm is to scan a sequence from left to right, and add new substrings to a vocabulary. Once the scan is complete, the

vocabulary size is the complexity [98]. The following describes the process of the Lempel-Ziv algorithm.

Let A be the binary alphabet where Λ represents the null character. We have A^* representing all possible binary strings and $\mathbf{s} \in A^*$ representing a single binary string of length $\ell(\mathbf{s})$. Each string may be specified by $\mathbf{s} = s_0, s_1, \ldots s_n$ where $n = \ell(\mathbf{s})$. If all the characters in a sequence are the same then $\mathbf{s}$ can be represented by s^n, which means to repeat character s, n times. Also, the notation $s(i, j) = s_i, s_{i+1}, \ldots, s_j$ will be used to denote a substring of $\mathbf{s}$ from position i to position j, where $i \leq j \leq n$. If $j > i$ then $s(i, j) = \Lambda$.

Consider a string $\mathbf{r} \in A^*$, where we want to measure its complexity. To do this, we must construct our vocabulary of words; i.e., substrings of $\mathbf{r}$. The first step in the process is to create a sliding window of variable size over $\mathbf{r}$. Let $\mathbf{q} \in A^*$ be this window where we start with $\mathbf{q} = r(j, i)$ with $i = 0$ and $j = 0$. Also, let $\mathbf{s} \in A^*$ be the concatenation of each word in our dictionary in the order that each word was added. When there are no words in the vocabulary, $\mathbf{s} = \Lambda$.

With $\mathbf{r}$, $\mathbf{q}$, and $\mathbf{s}$ defined, we now begin an iterative process asking the question: Can $\mathbf{s}$ generate the substring $\mathbf{q}$? If the answer is no, then we add $\mathbf{q}$ to the vocabulary while incrementing i and setting $j = i$, but if the answer is yes, then we let the window grow by one character by incrementing i by one. The generation of $\mathbf{q}$ by $\mathbf{s}$ follows the copy procedure described by Lempel and Ziv [98] in 1976. Essentially, the copy procedure copies characters in $\mathbf{s}$ to form the string $\mathbf{sq}$, where $\mathbf{q}$ is concatenated to $\mathbf{s}$. However, the key idea is that part of $\mathbf{q}$ may be assimilated into $\mathbf{s}$ in order to generate the rest of $\mathbf{q}$. Consider the following three cases to better understand this notion in which $\mathbf{s}$ is said to generate $\mathbf{q}$, when any of them hold.

1. If we are able to pick positions p and q where $p \leq q \leq \ell(\mathbf{s})$, then $\mathbf{q} = s(p, q) = s_p, s_{p+1}, \ldots, s_q$ and $\mathbf{s}$ can generate $\mathbf{q}$.
2. If each $q_k \in \mathbf{q}$ where $0 \leq k \leq \ell(\mathbf{q})$ is equal to the last character in $\mathbf{s}$, i.e., $s_{\ell(\mathbf{s})}$. This means that we can copy the last character of $\mathbf{s}$ as the first character of $\mathbf{q}$ and then the copy the first character of $\mathbf{q}$ to be the second character of $\mathbf{q}$, and so forth until $\mathbf{q}$ is generated. So $\mathbf{q} = s_n^m$ where $n = \ell(\mathbf{s})$ and $m = \ell(\mathbf{q})$.
3. If we can pick a positions $p < \ell(\mathbf{s})$ and $q = \ell(\mathbf{s})$, such that $\mathbf{q} = s(p, q)s_q^m$ where $m = \ell(\mathbf{q}) - q + p$, then we can generate $\mathbf{q}$ from $\mathbf{s}$. Note this case handles Cases 1 and 2, but all three are presented for clarity.

In the following example, let $\mathbf{r} = 0001101001000101$ be a cyclic string, $\mathbf{s} = \Lambda$, and $\mathbf{q} = r(j, i)$ where $i = 0$ and $j = 0$. The first iteration examines the window $\mathbf{q} = 0$. Since $\mathbf{s}$ is empty, we cannot generate $\mathbf{q}$ from $\mathbf{s}$, so let's add $\mathbf{q}$ to the vocabulary. Now $i = 1$ and $j = 1$ where $\mathbf{q} = r(1, 1) = 0$. This presents Case 1 and so $\mathbf{q}$ is the same as substring

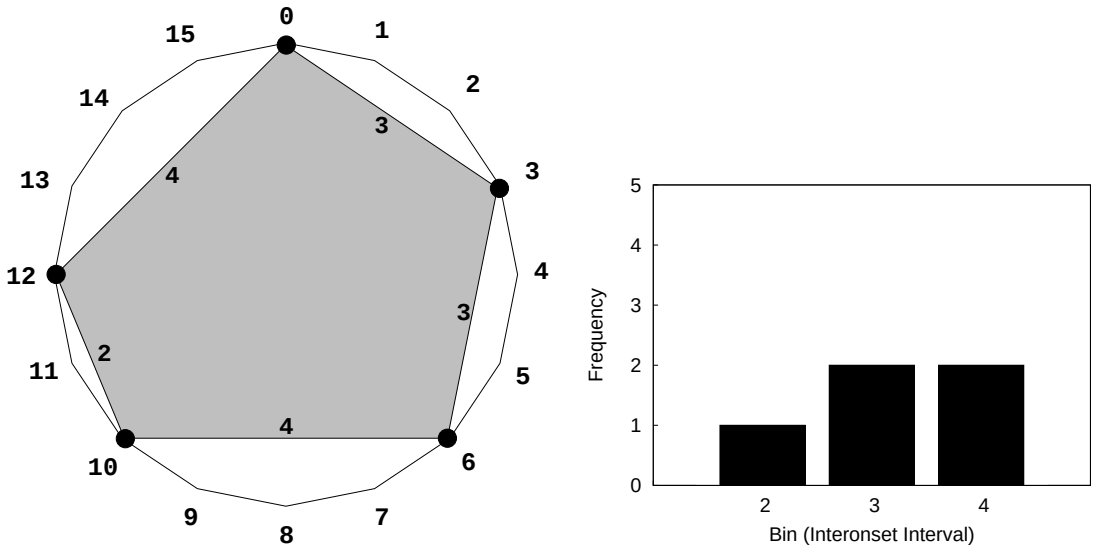
$s(p, q)$ where $p = 0$ and $q = 0$. Since $\mathbf{s}$ generates $\mathbf{q}$ we expand our window and set $i = 2$ and $j = 1$, thus $\mathbf{q} = r(1, 2) = 00$. Here we see that Case 2 applies so that s_n, where $n = \ell(\mathbf{s})$, is copied to q_0 and q_0 is copied to q_1. So we have $\mathbf{s}$ generating $\mathbf{q}$. Again we can expand the window and so $i = 3$ and $j = 1$. Hence $\mathbf{q} = r(3, 1) = 001$ and since the string representation of our vocabulary is $\mathbf{s} = 0$, there is no way we can generate 001. Thus we must add $\mathbf{q}$ to the vocabulary where we have $0 \cdot 001$. At this stage, our vocabulary size is 2, which is the complexity so far. The finished vocabulary of $\mathbf{r}$ requires two cycles of $\mathbf{r}$ to complete. The vocabulary is $0 \cdot 001 \cdot 10 \cdot 100 \cdot 1000 \cdot 101000$ of size 6. The *Lempel-Ziv Complexity* has been implemented, and is abbreviated as *lz*.

3.5 Interonset Interval Histograms

Histograms have a long history in the field of statistics as an approach for estimating the density of an unknown probability distribution[148]. The construction of a histogram relies on a set of bins of a determined width, where each bin contains a certain number of elements. The result is a distribution where the probability of a new element falling into a bin is determined by the number of elements in a bin divided by the total number of elements in all bins [148]. Here, a similar histogram approach will be used to create a distribution of the frequency of the intervals between onsets (i.e., durations of the onsets) in a rhythmic pattern. In the literature [33, 147, 172, 173], different approaches for generating the interonset intervals (IOIs) histograms have been seen. Note that an IOI is computed by counting the number of pulses between adjacent onsets. Below we discuss two methods: histograms based on the frequencies of *local* IOIs and histograms based on the frequencies of *global* IOIs.

First, consider histograms based on the frequencies of local IOIs. Measuring the IOIs locally (or relatively) consists of measuring the duration between sequential pairs of onsets in a rhythm and then counting the number of times each duration occurs. In the 1970s, Regener [131], Lewin [100], and Clough [22] used the term *interval normal form* to represent the directed distances between successive musical pitch classes in a set. In the 1980s, Clough and Myerson [23] again used this form in their work relating classes of musical chords. The interval normal form has additionally been used for rhythms. Coyle and Shmulevich [33] and Shmulevich et al. [147] describe a difference rhythm vector which is a vector of ratios between the intervals of sequential pairs of onsets. This form has been used as a measure for rhythmic similarity [33, 147, 173].

However, in order to generate the local IOI histograms we use the IOIs of sequential onsets. Thus, each bin of the histogram is a count of the number of times an IOI occurs. This construction is considered a frequency histogram of local IOIs. In order to visualize this, consider Figure 3.10, which shows the a polygon of the *clave son* rhythm along with



(a) Polygon. (b) Histogram.

Figure 3.10: *In (a) a polygon of the* clave son *rhythm,* $[\mathtt{x}\,.\,.\,\mathtt{x}\,.\,.\,\mathtt{x}\,.\,.\,.\,\mathtt{x}\,.\,\mathtt{x}\,.\,.\,.]$*, with local IOIs,* $(3, 3, 4, 2, 4)$*, is shown. These durations are determined by counting the pulses around the circle between adjacent onsets. The frequency of each IOI is captured in the histogram depicted in (b).*

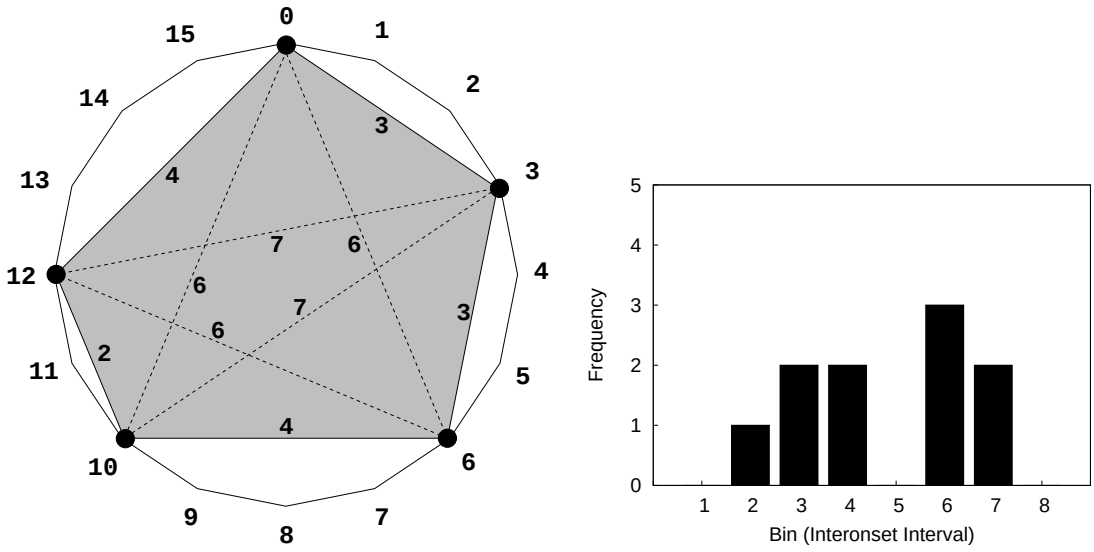
the corresponding histogram of the frequencies of the IOIs. Each IOI may be determined by counting the number of pulses around the circle between adjacent onsets in a clockwise fashion. Each edge length of the polygon is labeled with the IOI.

Second, consider histograms based on the frequencies of global IOIs. Measuring the IOIs globally consists of measuring the duration between all unique pairs of onsets in a rhythm. Block and Douthett [9] and McCartin [107] measure intervals between all pairs of pitch classes in a chord. Toussaint [172] has applied their methods to the rhythmic domain.

Given a cyclic rhythm with n pulses and k onsets, we have $\binom{k}{2}$ unique interval pairs. For each interval pair, we find the IOI between the onsets. A clear way to visualize this is geometrically. Consider Figure 3.11 which shows an example of this method for the *clave son* rhythm.

Once the rhythm is place on a circle, as seen in Figure 3.11(a), one may connect vertices of each unique onset pair. This is depicted by the solid lines which form the polygon, and also by the dotted lines which are all of the polygon diagonals; i.e., lines which connect pairs of non-adjacent vertices [183]. Now the distance of each edge may be measured in a geodesic manner (shortest path) by counting the pulses between onsets around the perimeter of the circle. Figure 3.11(b) captures the frequencies of the global IOIs.

Both ways of constructing IOI histograms may be used to measure the complexity of a musical rhythm. Below we discuss three methods. The first measures the standard deviation



(a) Polygon. (b) Histogram.

Figure 3.11: *In (a) a polygon of the* clave son *rhythm,* [x..x..x...x.x...]*, with global IOIs,* $(3, 3, 4, 2, 4, 7, 6, 7, 6, 6)$*, is shown. These durations are determined by counting the pulses (in a geodesic fashion) around the circle between all unique pairs of onsets. The frequency of each IOI is captured in the histogram depicted in (b).*

of the bin heights, the second measures the information entropy of the bins after they are normalized such that the sum of the frequencies is 1, and the third measure is based on the height of the tallest bin, after bin normalization. Consider the following descriptions of each complexity measure which may be applied to both techniques of histogram construction.

3.5.1 Standard Deviation

In order to measure the complexity of a rhythm's IOI histogram, the standard deviation is calculated. The standard deviation is the square root of the average squared deviations for some set of values from their mean [111]. Below we show the formula for the standard deviation, which may be found in any statistics textbook [111]; note that sd denotes the standard deviation, X is some set of values, $\overline{X}$ is the mean, and n is the number of values.

$$sd = \sqrt{\frac{\sum_{x \in X}(x - \overline{X})^2}{n - 1}} \tag{3.32}$$

The histograms present a set of values in which we may calculate the standard deviation. Therefore, Equation (3.32) is computed where X is the set of frequencies for each bin. However, we must take into account all possible bins. Thus when using the method for local IOIs, if a rhythm with 16 pulses has one onset at pulse zero, and the rest of the pulses are silences, then the histogram would have bin 16 with a frequency of 1. So, for each histogram there are 16 possible bins. In the case of the *clave son*, we have frequencies

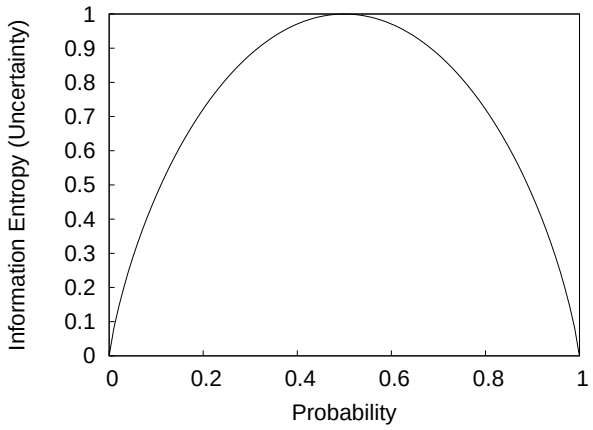


Figure 3.12: *Information entropy (uncertainty) in the case of two possible values where the probability varies, note that the peak is when the probability of obtaining both values is* 0.5.

for bins two, three, and four, where the rest of the bins would have zero frequencies. This method was also chosen for the global IOI histogram standard deviation calculation, which is opposed to using a value of zero when a rhythm with a single onset occurs. Musically, a duration of zero would mean that nothing has sounded, but here we do have an onset and so something has sounded. However, during testing, both methods were found to be very similar on the data used in this thesis. When a rhythm is found to have a low standard deviation, the histogram is very even (i.e., flat), which means that a variety of IOIs are present in the rhythm. Thus, we consider this to mean that the rhythm is more complex. Using the standard deviation, we have two measures for complexity which are abbreviated as: *ioi-l-sd* for the local IOI histogram, and *ioi-g-sd* for the global IOI histogram.

3.5.2 Information Entropy

Information entropy may also be used to measure the complexity of a rhythm from IOI histograms. If we normalize the bins in the histogram such that the sum of the frequencies is equal to 1, then such a normalized histogram may be considered a discrete probability distribution. Consider $\mathcal{X}$ to be a discrete random variable which takes on the IOI of each bin with probability corresponding to the normalized frequency. We may calculate the Shannon information entropy [143], also referred to as the uncertainty [64, 141, 179], of $\mathcal{X}$, by the formula for information entropy, previously seen in Equation (3.16). The uncertainty will be at its maximum when the probability distribution is uniform [141]. This can be seen in the binary case of two possible values, by Figure 3.12.

Figure 3.12 shows that the uncertainty is highest when the probability distribution is uniform in the case of two possible values. When there are more than two values, this also holds, the uncertainty of a random variable is at its maximum in the case of a uniform

underlying probability distribution [141]. Thus, we may use the measure of information entropy to indicate the uniformity, or the flatness, of the normalized IOI histograms. The more flat the IOI histogram, the more complex the rhythm, because our assumption is that a rhythm, which admits a greater variety of onset durations, is more complex.

In our implementation we calculate the uncertainty as follows. Assume we have previously computed an IOI histogram for a given rhythm. Let X be the set of IOIs, where each value in X represents a bin label of the histogram. Define a function $f : X \rightarrow \mathbb{Z}$, which returns the frequency count for a given IOI based on the histogram. Note that $\mathbb{Z}$ is the set of integers. Let c be the sum of all values $f(x)$ where $x \in X$. Now we may define a probability mass function p using f and c.

$$p(x) = \begin{cases} \frac{f(x)}{c} & \text{when } x \in X \\ 0 & \text{otherwise} \end{cases} \tag{3.33}$$

Let $\mathcal{X}$ be a discrete random variable which takes on values from X with probability mass function p. We now calculate the uncertainty of $\mathcal{X}$ as follows.

$$H(\mathcal{X}) = -\sum_{x \in X} p(x)\ log_2\ p(x) \tag{3.34}$$

The value $H(\mathcal{X})$ is the uncertainty that $\mathcal{X}$ will take on IOI values, which are based on probabilities from the mass function p. $H(\mathcal{X})$ will increase when the distribution is more uniform. This means that the rhythm (from which the distribution was derived), will be more complex. The uncertainty is calculated in this manner using both local IOI histograms and global IOI histograms, which yields two measures: *ioi-l-h* for local IOI histograms and *ioi-g-h* for global IOI histograms.

3.5.3 Tallest Bin

Another way to measure the flatness of the histogram to determine a rhythm's complexity is by looking at the height of the tallest bin. To do this, we follow the same normalization procedure of creating a probability mass function described in §§ 3.5.2. Thus, assume that p exists based on an IOI histogram of a given rhythm. Moreover, let X be the set of IOIs, bin labels, of the histogram. Find the maximum value of $p(x)$ over each $x \in X$.

$$\max_{x \in X} p(x) \tag{3.35}$$

This value tells us the height of the tallest bin. If the tallest bin has a small height, this means that the histogram is more uniform. Recall that we assume that rhythms, which are more complex, have more uniform IOI histograms. Thus, we use this approach for both

local IOI histograms and global IOI histograms to measure the complexity of a rhythm. The abbreviations for this measure are: *ioi-l-mm* for local IOI histograms and *ioi-g-mm* for global IOI histograms.

3.6 Mathematical Irregularity

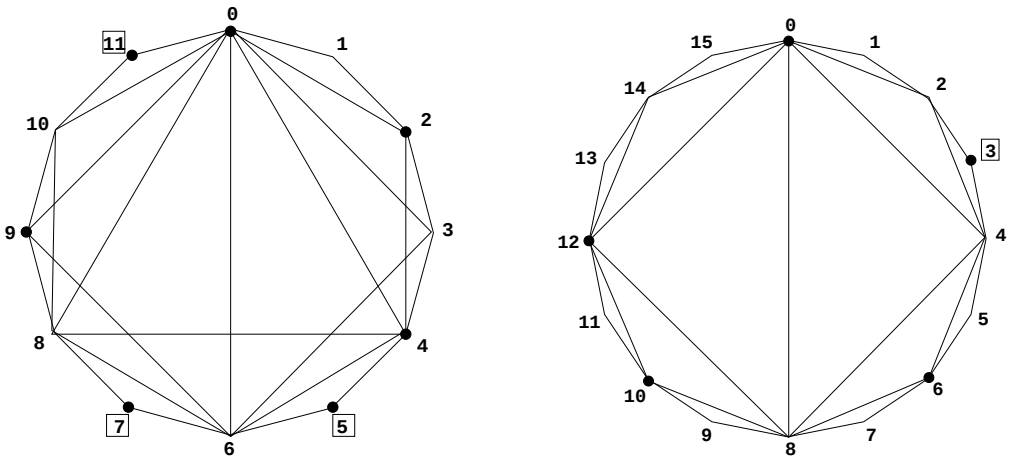
Finally, complexity measures are presented which pertain to measuring how irregular a rhythm is, by way of a geometric representation. Both measures have been discussed by Toussaint [168, 169].

3.6.1 Toussaint's Off-Beatness

The *Off-Beatness Measure* proposed by Toussaint [67, 68, 169], can be used as a measure for rhythmic complexity. In order to describe the Off-Beatness of a rhythm, consider a geometric representation. Let $\mathbf{r}$ be a cyclic rhythm with n pulses and k onsets. First, place $\mathbf{r}$ on a circle of fixed radius where each pulse is evenly distributed around the circumference of the circle. Second, find each value i that is greater than 1, less than n, and which evenly divides n. Third, for each i, inscribe a regular polygon with i vertices on the circle, starting the inscription at pulse 0 and continuing clockwise [67, 169]. Fourth, mark each pulse which corresponds to a vertex of an inscribed polygon a beat, and mark each pulse which does not correspond to a vertex of any inscribed polygon an *off-beat*. Fifth, count the number of onsets which occur on an off-beat [67, 169]. This value is the Off-Beatness measure of $\mathbf{r}$. Since we are looking for onsets around the circle which lie on a vertex not shared by a *regular* inscribed polygon, we say this is a measure of *irregularity*.

As an example of Off-Beatness, let $\mathbf{r}$ be the *bembé* rhythm, [x.x.xx.x.x.x], where $n = 12$ and $k = 7$. First, let us place $\mathbf{r}$ on a circle starting with pulse 0 numbering clockwise to pulse 11. Second, the numbers in the range $(1, 12)$ which evenly divide 12 are: $2, 3, 4$, and 6. For the third and fourth step, if we inscribe a regular bi-angle, triangle, square, and hexagon, where we start drawing the polygon at pulse 0 and continuing clockwise, then the beats (polygon vertices) fall on pulses $0, 2, 3, 4, 6, 8, 9$, and 10. The off-beats (non-polygon vertices) are thus $1, 5, 7$, and 11. Counting the onsets which occur on an off-beat, we arrive at the off-beatness for $\mathbf{r}$ to be 3. Figure 3.13 visualizes the example, and presents an example on a 16-pulse rhythm. Note that we abbreviate this measure as *offbeatness*.

Off-Beatness also has group and number theoretic implications [67], which we briefly touch on here. In group theory, let G be a group of the integers in the range $[0, n-1]$ under the operator of addition modulo n. In other words, let G be $\mathbb{Z}/n\mathbb{Z}$. Let S be a subset of G, such that for each $s \in S$, we have s and n are relatively prime. Each element in S generates G and thus $\langle S \rangle = G$. Additionally, in number theory, the order of S is precisely the result of Euler's totient function.



(a) *Bembé* rhythm. (b) *Clave son* rhythm.

Figure 3.13: *The Off-Beatness measure applied to the* bembé *rhythm in (a) as an example of a rhythm with 12 pulses, and in (b) as an example of a rhythm with 16 pulses, the* clave-son*. Note that the off-beat onsets are marked with the corresponding pulse in a box.*

3.6.2 Rhythmic Oddity

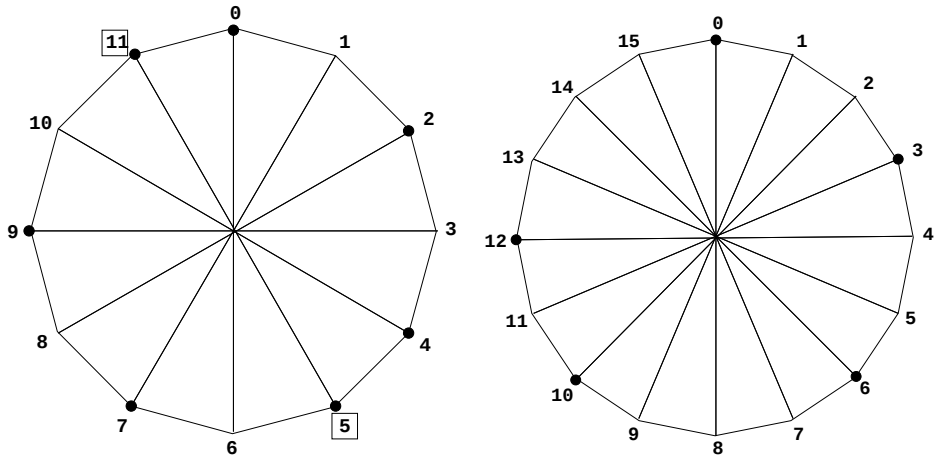
The *rhythmic oddity property* was discovered by Simha Arom [4] and used to describe a property that is characteristic of rhythms of the Aka Pygmies [17, 67, 169]. A rhythm is said to have the rhythmic oddity property if there is no pair of onsets that partition the rhythm into two sub-units with an equal number of pulses [67, 169]. Implied in this definition is that the rhythm is cyclic, and also the rhythmic oddity property is only relevant to rhythms with an even number of pulses, since an odd number of pulses implies the property [67].

In order to better understand the definition of the rhythmic oddity property, consider a rhythm $\mathbf{r}$ with n pulses and k onsets. Place rhythm $\mathbf{r}$ on the perimeter of a circle of n evenly spaced points, such that each pulse corresponds to a point. If a diameter of the circle cannot be drawn between any onset pair, then $\mathbf{r}$ has the rhythmic oddity property [17].

Rhythmic oddity can also be used to measure complexity. Toussaint [67, 169] has developed an algorithm to measure the rhythmic oddity of a rhythm, called the *Rhythmic Oddity* measure. Let $\mathbf{r}$ be a cyclic rhythm with an even number of n pulses and some k number of onsets, where $\mathbf{r}$ is indexed from 0 to $n-1$. The first step of the algorithm is to iterate through all pairs of onsets $(i, i+n/2)$ where $0 \leq i < n/2$. The second step is to check if r_i and $r_{i+n/2}$ are both onsets. If so, then a partition with two sub-units having an equal number of pulses can be created. Thus, add a 1 to total, which has been initialized to 0. Once all such pairs have been checked, the total is the number of onset pairs which can partition the rhythm. The smaller the number of onset pairs, the greater rhythmic oddity (closer to having the rhythmic oddity property) and thus the greater the complexity of the

rhythm [169]. This also presents a measure of irregularity because a diameter would split the number of pulses equally in half. However, we are looking for irregular pairs of onsets which do not equally partition the number of pulses.

As an example, consider the *bembé* rhythm where $\mathbf{r} = [\texttt{x.x.xx.x.x.x}]$, $n = 12$ and $k = 7$, where we index $\mathbf{r}$ from 0 to 11. Let us now check the pairs: $(0, 6)$, $(1, 7)$, $(2, 8)$, $(3, 9)$, $(4, 10)$, and $(5, 11)$. We see that one pair, $(5, 11)$, has onsets at each position in $\mathbf{r}$, and hence, creates a partition where the number of pulses are equal. Thus the rhythmic oddity is 1 for the *bembé* rhythm. Figure 3.14 provides and example for the *bembé* and *clave son* rhythms. The *clave son* admits no diameters. Note that the abbreviation for this measure is *oddity*.



(a) *Bembé* rhythm. (b) *Clave son* rhythm.

Figure 3.14: *The Rhythmic Oddity measure applied to the* bembé *rhythm in (a) as an example of a rhythm with 12 pulses, and (b) presents and example of the Rhythmic Oddity applied to a rhythm with 16 pulses, the* clave-son*. Note that the onset pair which creates a diameter are marked with the corresponding pulses in a boxes.*

Chapter 4

Experimental Data

The experimental rhythm data used in the evaluation of the complexity measures described in Chapter 3, shall be discussed. There are three main categories which organize the rhythms. The first is termed *Psychological*, as the rhythms in this category have been used in psychological experimentation. Here, human measures of rhythm complexity are also described, which have been collected from four different psychological studies. The second category is labeled *Cultural*, as the rhythms in this category have been gathered from two different cultures: African and Indian. The third category is labeled as *Random*, because the rhythms here have been generated at random. Below, each category of rhythms will be described and tables of the rhythms are included.

4.1 Psychological

The rhythms in this section have been used in four different psychological studies. However, two studies use the same set of rhythms, and thus we present three rhythm data sets. The subsections are grouped by rhythm data set, and so first we present the rhythms originating from a psychological study by Povel and Essens [119] in 1985. In this subsection, we discuss the experimentation by Povel and Essens on human participants with the rhythms, and also discuss a second study, conducted by Shmulevich and Povel [146] in 2000, where the data set is used as well. Povel and Essens' study yields a human measure of rhythm performance complexity, whereas Shmulevich and Povel's study yields a human measure of rhythm perceptual complexity regarding the rhythms listed in Table 4.1. The second rhythm data set originated from a psychological study by Essens [48] in 1995. As a result of this study, human measures of rhythm performance complexity and rhythm perceptual complexity are also obtained. The rhythms from Essens' study are listed in Table 4.2. The third rhythm data set is the most recent data set, from a study by Fitch and Rosenfeld [57] in 2007. From this study, we obtain a human measure of rhythm performance complexity and also two human measures of rhythm metrical complexity. The rhythms are listed in Table 4.3.

4.1.1 Povel and Essens; Shmulevich and Povel

In 1985, Povel and Essens [119], conducted three experiments to gain insight regarding the internal representation of rhythms. In 2000, Shmulevich and Povel [146] conducted an experiment in order to gain "perceptual validity" [146] of three measures of rhythm complexity [146]. Both studies used the data shown in Table 4.1. The results from Povel and Essens' Experiment 1 are used by this thesis as a baseline pertaining to the performance

complexity of the rhythms. The results from Shmulevich and Povel's experiment are used as a baseline pertaining to the perceptual complexity of the rhythms. This subsection is divided as follows: first, we list the rhythms along with the human-based complexity measures in Table 4.1. Then we describe the method Povel and Essens used to obtain, what we call, *Povel and Essens' Human Performance Complexity* measure. Finally, we discuss the method used by Shmulevich and Povel to obtain, what we term, *Shmulevich and Povel's Human Perceptual Complexity* measure.

Povel and Essens' Experiment

The experiment by Povel and Essens [119] used synthetic rhythms (see Table 4.1) to test the hypothesis, "temporal patterns [rhythms] that strongly induce an internal clock [metrical representation] form better internal representations and are consequently better reproduced than weaker clock-inducing patterns" [119]. These rhythms are labeled as synthetic because they are all permutations of the IOI pattern $\{1, 1, 1, 1, 1, 2, 2, 3, 4\}$, which admits only structural variations according to Povel and Essens [119].

The experiment to test their hypothesis is relevant to rhythmic complexity, because Povel's clock-model is related to the how humans form a metrical hierarchy when given a rhythm, predicting its structure [120, 121]. Therefore, in Povel and Essens' study, a method of rhythm reproduction accuracy is used to test if a rhythm, which presents a clearer metrical structure, can be more accurately reproduced [119].

Twenty-four subjects participated in the experiment where a rhythm was sounded and then stopped, to allow the participant to tap back the rhythm just heard [119]. The rhythm could be listened to for as long as the participant desired, and subjects were also encouraged to tap along to the rhythm as it played. However, once they pressed the stop button, the participants were asked to tap back the rhythm repeatedly four times. If the participants were dissatisfied, they could repeat the process for the rhythm just heard. Otherwise, the next rhythm was randomly presented [119].

After the experimentation, the reproduction accuracy was measured in terms of what Povel and Essens call a *mean deviation percentage* [48, 119]. This value was calculated by first finding the IOI values between all reproductions from the participants and from computer sounded target rhythms. Second, the average over the four reproductions of each rhythm, from each participant, was calculated. Third, the percent difference was found between the average IOIs for the reproductions, and corresponding IOIs for the target rhythms, for each participant. Fourth, the average of the percent difference was found [48, 119]. This value is the mean deviation percentage.

This thesis interprets the mean deviation percentage values as ordinal ranks and refers to the values as the human-based performance complexity for the rhythms. Ranks are used

Table 4.1: *Rhythms originally from a study by Povel and Essens [119], which were also used in a study by Shmulevich and Povel [146]. Along with the rhythms, two measures of complexity of the rhythms are also shown, a higher value means the human participants found the rhythm more complex.*

No.	Rhythm	Human Performance Complexity	Human Perceptual Complexity
1	xxxxx..xx.x.x...	5	1.56
2	xxx.x.xxx..xx...	1	2.12
3	x.xxx.xxx..xx...	0	2.08
4	x.x.xxxxx..xx...	2	1.88
5	x..xx.x.xxxxx...	3	1.80
6	xxx.xxx.xx..x...	9	2.44
7	x.xxxx.xx..xx...	7	2.20
8	xx..xxxxx.x.x...	4	2.56
9	xx..x.xxx.xxx...	14	3.00
10	x.xxx.xxxx..x...	18	2.04
11	xxx.xx..xx.xx...	19	2.76
12	xx.xxxx.x..xx...	15	2.72
13	xx.xx.xxxx..x...	13	3.00
14	xx..xx.xx.xxx...	17	3.16
15	x..xxx.xxx.xx...	10	2.04
16	xx.xxxx.xx..x...	11	2.88
17	xx.xxx.xxx..x...	17	2.60
18	xx.xxx..xx.xx...	22	2.60
19	xx..xx.xxxx.x...	21	2.64
20	xx..xx.xxx.xx...	25	3.24
21	xxxxx.xx.x..x...	29	3.08
22	xxxx.x..xxx.x...	20	3.04
23	xxx..xx.xxx.x...	16	3.04
24	x.xxx..x.xxxx...	6	2.56
25	x.x..xxxx.xxx...	8	2.56
26	xxxx.x.x..xxx...	26	2.84
27	xx.xxx.x..xxx...	23	3.60
28	xx.x..xxx.xxx...	32	2.68
29	x.xxxx.x..xxx...	28	3.28
30	x..xxxxx.xx.x...	21	3.08
31	xxxx.xxx..x.x...	30	3.52
32	xxxx..xx.xx.x...	31	3.60
33	xx.xxxx..xx.x...	24	3.04
34	xx.x..xxxxx.x...	33	2.88
35	x.x..xxx.xxxx...	12	3.08

since the exact values were unavailable, and so to avoid possible error, ordinal ranks were obtained from Povel and Essens' work [119]. In Table 4.1, the ordinal ranks can be seen in the last column. A higher rank indicates that rhythm produced a higher mean deviation percentage, and thus a higher complexity.

Shmulevich and Povel's Experiment

Now consider Shmulevich and Povel's [146] experimental method for obtaining the human judgments of perceptual complexity shown in Table 4.1. Twenty-five subjects participated in their study. The participants consisted of a mixture of graduate, undergraduate, and faculty members from the University of Nijmegen. All the participants were musicians with an average of 9.2 years of musical experience [146].

Each rhythm was sounded by a MIDI-synthesizer generating a marimba sound. Each onset in the rhythm sounded the same, there were no accents or changes in pitch. Each rhythm was played four times in a row, where the next rhythm was chosen at random [146]. This presentation was possible because Shmulevich and Povel used a computer program to display, sound, and collect the user-responses for each rhythm. The responses were in terms of a judgment on the complexity of the rhythm, where the participant could give integer values starting at 1, meaning simple, up to and including a 5, meaning complex. The instruction statement given to the participants was, "imagine how difficult it would be to reproduce the rhythms" [146]. Hence, the participants were given a performance mindset.

The average among the resulting human complexity judgments for each rhythm are presented by Shmulevich and Povel [146], and are also shown in Table 4.1, labeled *human perceptual complexity*. As a note on the average complexities, Shmulevich and Povel tested the consistency with Cronbach's alpha, a statistical test which shows reliability in the complexity values [34].

4.1.2 Essens

In 1995, Essens [48] conducted three experiments which were focused on the structure of rhythms. This thesis adopts the results from Experiment 2 and Experiment 3 of Essens' work [48]. The results are used as another baseline measure for human-based performance and perceptual complexity, but on a new data set of rhythms listed in Table 4.2.

The performance complexity is from Experiment 2 conducted by Essens. Here, the reproduction accuracy of participants on the rhythms was measured again using the mean percent deviation, which is calculated between the IOIs in the reproduced rhythms and the IOIs of the computer-sounded target rhythms [48]. Perceptual complexity is derived from Essens' Experiment 3. Here, human judgments of the complexity for each rhythm is determined, again based on a 1 to 5 scale.

Table 4.2 provides the rhythms, along with the *Essens' Human Perceptual Complexity*

values, and *Essens' Human Performance Complexity* values as ordinal ranks. The reason ordinal ranks for the performance complexity are used is again because exact values were not available. Therefore, rankings were determined from Essens' work [48]. Consider the table of each of the values, and then consider descriptions of Essens' Experiment 2 and 3.

Table 4.2: *Rhythms from Essens [48], along with two measures of human-based complexity for each rhythm from two experiments conducted by Essens [48].*

No.	Rhythm	Human Performance Complexity	Human Perceptual Complexity
1	xxx.xxx.xxx.xx..	0	2.2
2	xxx.x.xxx.xxxx..	8	3.1
3	x.xxx.xxx..xxx..	4	3.2
4	x.xxx..xx.xxxx..	19	2.9
5	xxx.xxx.xx.xx...	2	2.2
6	xxx.x..xx.x.xx..	7	3.1
7	xxxxxxx.xxx.xxx.	10	2.6
8	xxx.xxxxx..xxx..	5	4.2
9	xxxxxx.xx.xxx...	13	2.9
10	x.x.x.x.xxx.xx..	6	2.8
11	xxxxxxx.xxx.x.x.	1	3.1
12	xxx.xx..x.x.x...	3	2.5
13	x..xxxx.xx..xx..	20	3.5
14	x.xxxx.xxx.xxx..	12	2.5
15	x..xxx.xxx.xxx..	14	2.4
16	x..xxx.xxxx.xx..	11	3.0
17	xx.xxx.xxxx.x...	17	3.0
18	x..xxxxxxx.xxx..	18	3.1
19	x.x.xx.xxx.xxx..	22	2.4
20	xx.xxxx.xx.xx...	16	3.2
21	xx.xxx.xxxxxx...	15	2.4
22	xx..xx.xxxxxx...	11	2.9
23	x.x.xx.xxxxxxx..	21	2.7
24	xx.xxxxxx.x.xx..	9	3.8

Experiment 2

Experiment 2 conducted by Essens involved six participants, where two of them played a musical instrument for at least 5 years [48]. Each rhythm, as shown in Table 4.2, was generated to vary how well the rhythms define a metrical structural based on Povel's clock-model [119, 120, 121].

Essens presented each rhythm cyclically four times. Participants were asked to listen to each rhythm and tap along. Then, they stopped the computer generated rhythm, and had to reproduce it. This was done by tapping the rhythm as accurately as possible four times [48]. This was the immediate reproduction task, where the results are shown in Table 4.2 in the *human performance complexity* column. These data are the ranks of the mean deviation percent from the top panel of Essens' Figure 2 [48]. The mean deviation percentage was

calculated in the same way as in Povel and Essnens' study [119]. Thus, the higher the mean deviation percentage, the more difficult it was to reproduce the rhythm, and thus the higher the ranking of that rhythm.

Experiment 3

Experiment 3 conducted by Essens [48] had twenty participants, where five were musically trained for at least 5 years. The purpose of this experiment was to judge complexity of the rhythms. For each rhythm, the participants judged the complexity on a five-point scale: very simple, moderately simple, neutral, moderately complex, very complex [48]. The rhythms were randomly ordered, and there was no definition of complexity given. These values are listed in Table 4.2, labeled *human perceptual complexity*. They represent the mean complexity over the participants for each rhythm.

4.1.3 Fitch and Rosenfeld

Most recently, Fitch and Rosenfeld [57] conducted an experimental study of rhythm reproduction (i.e., performance complexity) and beat-tracking (i.e., rhythmic meter complexity). The rhythms used for their study are shown in Table 4.3. These rhythms were generated in such a way as to vary the amount of syncopation among the rhythms, as measured by Fitch and Rosenfeld's implementation [56] of the Longuet-Higgins and Lee measure [106]. The 30 rhythms obtained, used in both the reproduction task and beat-tracking task, where 16 participants with musical experience ranging from 0 to 15 years (and 6 of them seriously involved in regular musical activities), were tested [57]. From both tasks, we derive human-based complexity measures. Fitch graciously shared [55] data used here.

Reproduction Task

The experiment by Fitch and Rosenfeld for measuring the accuracy of rhythm reproduction involved asking the participants to reproduce (by tapping) the rhythm they had just heard. That is, the subject must tap the rhythm, as the computer taps the beat of the rhythm [57]. So, this experiment was combined with the beat-tracking experiment (to be described next), in such a way that after the participant heard the computer play the rhythm for the beat-tracking task, the roles switched. Here in the reproduction task, the participant reproduces the rhythm just heard, while the computer plays the beat [57]. To measure the rhythm reproduction accuracy, Fitch and Rosenfeld took the IOIs of both the target rhythm sounded by the computer, and the IOIs from the reproduced rhythm by the participant. Then they optimally aligned the IOIs of the reproduced rhythm to the target rhythm's IOIs, and computed the absolute value of the difference for each IOI. These error values where then averaged for each IOI pair and then an average across the participants was taken for the reproduction error for a rhythm [57]. We consider these values to be *Fitch and Rosenfeld's Human Performance Complexity*, as shown in Table 4.3.

Beat-Tracking Task

In the beat-tracking experiment, a computer sounds a steady beat (which Fitch and Rosenfeld call pulse), and then sounds a rhythm simultaneously along with the beat. Participants were then asked to tap along with the beat and try to accurately follow the computer's beat. After two set periods of the beat and then two periods of the beat and rhythm together, the computer-generated beat stops; however, the participant must keep tapping (i.e., maintain) the beat [57]. The tempo was varied during this process, and the result was termed the *beat tapping error*. Fitch and Rosenfeld also have a measure for the beat tapping error adjusted for the tempo variations. We call this *Fitch and Rosenfeld's Human Metrical Complexity (Beat Tapping Adj)*, as seen in Table 4.3.

An additional measure of human-based metrical complexity of was collected from the beat-tracking experiment. This is termed as the *number of resets* [57]. Here, we use the number of resents for *Fitch and Rosenfeld's Human Metrical Complexity (Number of Resets)* measure, as shown in Table 4.3.

Fitch and Rosenfeld describe a *reset* as an event when listeners alter their notion of where the beats are in a rhythm; i.e., "they 'rehear' or reinterpret the pattern as less syncopated by shifting the inferred pulse [beat]" [57]. This was tested by counting the number of times a participant tapped a beat half-way between where the beat should have been [57].

4.2 Cultural

The rhythms presented in this section consist of those rhythms from two different cultures: African and Indian. First we discuss the rhythms from African culture, termed African *Timelines*, and then we discuss the Indian rhythm. However, we have two data sets of rhythms from Indian culture: the Indian *Decitalas* and the North Indian *Talas*.

4.2.1 African *Timelines*

The famous musicologist A. M. Jones writes that, "an interesting rhythmic pattern is the spice of life to the African" [82]. Since rhythm is so characteristic in African culture [82], we include a handful of rhythms collected from a variety of sources [1, 2, 181, 82, 124, 128, 165, 170, 171, 174] in this study. However, most notably are the sources by Toussaint [170, 171, 174], for providing such an extensive list of African *Timelines*.

The rhythms used have either 12 pulses or 16 pulses, and in fact are a collection of African *derived* rhythms. Hence, Table 4.4 shows mostly African bell or clapping patterns, but there are rhythms in the list which are considered to be derived or influenced from Africa. For example, the flamenco clapping patterns *guajira*, *seguiriya*, and *soleá*, are either identical to or rotations of African rhythms, and *bulería* is a close relative to *soleá*. Such patterns are thought to perhaps have originated in north Africa, later spreading to Spain [85]. We include the *macedonian* rhythm because it is simply a rotation of the *guajira*

Table 4.3: *Rhythms from Fitch and Rosenfeld's [57] study, along with a human-based measure of performance complexity and two human-based measures of metrical complexity.*

No.	Rhythm	Human Performance Complexity	Human Metrical Complexity (Beat Tapping Adj)	Human Metrical Complexity (Number of Resets)
1	x.....x.x.....x.	0.138	0.075	2.500
2	x...x...x...x.x.	0.145	0.082	2.250
3	x.x.x.......x.x.	0.153	0.075	2.313
4	x...xxx.....x...	0.257	0.119	8.750
5	..x...x.x.x.x...	0.133	0.103	5.500
6	..x...x.x...x.x.	0.235	0.082	3.063
7	x.....x..x....x.	0.215	0.112	6.000
8	x..x...x..x.x...	0.208	0.110	5.188
9	x....x...x...xx.	0.250	0.141	6.938
10	x....x...x...x..	0.171	0.144	10.375
11	..x...x...xx.x..	0.220	0.130	6.875
12	.x...x...x..x.x.	0.226	0.124	6.438
13	..x...xx...x.x..	0.387	0.130	6.965
14	..x..x.....x...x	0.239	0.159	11.688
15	...x.x....xx...x	0.485	0.172	13.688
16	x...x.xxx.......	0.173	0.085	2.625
17	x.x.x.....x.x...	0.179	0.077	2.313
18	..x.x...x.x.x...	0.182	0.077	2.438
19	x.x...x.x..x....	0.252	0.074	1.938
20	..x...x.x...x...	0.142	0.098	3.375
21	x....x.x..x.x...	0.305	0.161	11.063
22	..x...x.x.x...x.	0.321	0.129	8.500
23	xx...x....x...x.	0.320	0.145	7.375
24	.x...x...xx.x...	0.265	0.134	7.188
25	..x...x...x..x..	0.176	0.146	8.625
26	..x...x...xx...x	0.326	0.118	6.500
27	..x..x....x..x.x	0.368	0.117	6.188
28	..xx...x...x.x..	0.344	0.154	10.813
29	.x.x.x.x...x....	0.185	0.191	15.750
30	.x.x...x...x...x	0.158	0.164	11.938

clapping pattern and thus, is closely related. Also, the *tuareg* from Lybia is included since it is from north Africa. The *persian* rhythm is also included in the list because of its relation to the African *sorsonet* by rotation; moreover, the Arab rhythm *al-ramal* is a rotation is three African rhythms (e.g. *tonada*) and is thus included as well.

Since African derived rhythms also includes rhythms stemming from Afro-American culture, rhythms from New Orleans such as the *tipitina* snare drum pattern is in the data set. In addition, rap-style rhythms and two other snare drum rhythms are used. Two other rhythms deemed *experimental*, have been included as well. The first, is a clapping pattern from Steve Reich [132], *steve-reich-clap*, which is closely related to the African *yoruba* rhythm [166, 167], being shown to in fact be a rotation [27]. The second, is a mathematical *cousin* to Reich's rhythm presented by Joel Haak [70, 71]. Haak provided a mathematical explanation to how Reich arrived at his rhythm. Among all possible rhythms with 12 pulses and 8 onsets (i.e., $\binom{12}{8}$), Haak devised a set of rules for picking-out clapping rhythms, and came up with only two: the rhythm termed *joel-haak* and Reich's rhythm. Thus, because of the mathematical properties these two rhythms share, *joel-haak* is also considered African derived, and included in our list. However, for simplicity, we use the term African *Timelines* to encompass all the rhythms listed in Table 4.4.

4.2.2 Indian *Decitalas*

The second set of rhythms is from Indian culture and are referred to as the Indian *Decitalas*. These rhythms are significant to Indian music as they are notated according to Ṡarṅgadeva, who wrote the thirteenth-century treatise, the *Saṅgīta-Ratnākara*, which is considered to be the "crucial text of Indian musical history" [122]. Moreover, the famous French composer Olivie Messiaen was greatly influenced by the *Decitalas*, as he discusses his rhythmic technique in his treatise *Technique de Mon Langage Musical* [109].

Thus, we use a subset of the *Decitalas* presented by Ṡarṅgadeva. The rhythms are from a text by R. S. Johnson [81] from 1975, where they've been replicated in Johnson's Appendix II, page 194. Listed by Johnson are 132 rhythms, which we reduce to 97, removing rhythms such as those containing only one onset, and those rhythms with a prime number of pulses, for example. The reason for removing those with a prime number of pulses, is because a majority of our complexity measures rely on prime factorizations of the number of pulses, and when that number is prime, we have little to work with. Therefore, those rhythms are removed and we present the reduced set Indian Decitalas in Table 4.5. Note that these Decitalas present a wide range of pulse values, so in the results, we take the full 97 rhythms, and also a subset taking only those rhythms with 12 or 16 pulses as an additional comparison to the African *Timelines*. Regardless, consider the Indian *Decitalas* below.

Table 4.4: *African* Timelines.

Name	Rhythm	Name	Rhythm
adangme	x.x.x.x.x..x.x..	kpatsa	x...x...x..x.x..
akan-five	x..x.xx.x...	kromanti	x.x...x...x.x...
akan-four	x..x..x.x...	mambo-a	x..x.x..x.x..x..
akom	x.x.x.x..x.x..x.	mambo-b	x.x..x..x..x.x..
al-ramal	x.x.x.x.xx.x	macedonian	x..x.x.x.x..
asaadua	x.x.x.xx.xx.	menjani	x.x.xx.xx.x.
ashanti-seven	x.xx.x.xx.x.	mutuashi	xx.x.x.xx.x.x.x.
ashanti-six	x.x.x..x.xx.	nandom-bawaa	x...x.....x.....
ashanti	x..x.x..x...	ngbaka	x.x.x.xx.x.x.xx.
beat-four	x..x..x..x..	ngbaka-maibo	x.x.x.xx.x.x.x.x
beat-six	x.x.x.x.x.x.	nyayito	x.xxxxxx.x.x
beat-three	x...x...x...	oyaa	x.x.x.x.x.xx.xx.
bemba-five	x..x.x.x..x.	persian	x.x.x.x.xxx.
bemba-seven	x.xx.x.x.xx.	popcorn-snare	x.x.x..x.x....x.
bembe-duple	x..x..xx..x.x..x	rap-drum	x...x..x.x..x...
bembe-five-a	x.x....x.x.x	rap-kick-drum-public-enemy	x.x....x..x...x.
bembe-five-b	x...x..x.x.x	red-white-blue-snare	x.x..x..x..x..x.
bembe-four-a	x.x....x.x..	rumba	x..x...x..x.x...
bembe-four-b	x...x..x.x..	rumba-palitos	x.xx.x.xx.x.xx.x
bembe-seven-a	x.x.xx.x.x.x	salve	x.x..x.x..x.
bembe-seven-b	xx.x.xx.x.x.	samba	x.x..x.x.x..x.x.
bembe-three	x.x.....x..	seguiriya-clap	x.x.x..x..x.
bondo	xx.x.x.xx.x.	shiko	x...x.x...x.x...
bossa-a	x..x..x...x..x..	sologon	x.xx.xx.xx.x
bossa-b	x..x..x..x..x...	solea-clap	..x..x.x.x.x
bossa-c	x..x..x..x...x..	son	x..x..x...x.x...
buleria-clap	..x...xx.x.x	sorsonet	xxx.x.x.x.x.
central-cuba	x.x.x...xx..	soukous-drum	x..x..x...xx....
columbia	x.x..x.x.x..	steve-reich-clap	xxx.xx.x.xx.
domba	x.x..x...x...x..	takoe	x..x.x..x.x.x.x.
dunumba	x.x.x.x.x..x	tambu	x.x.x.xx.x.x
eleggua	xxx.xx.x.x.x	timini	x.x...x.x...x...
equatorial-african-rattle	x.x.x.x.x..x..x.	tipitina-snare	x.x.x....x....x.
fume-fume	x.x.x..x.x..	tonada	x.xx.xx.x.x.
funky-drummer-snare	x.x.x..x.x..x...	tresillo-a	x.....x.....x...
ga	x.x.x.x..x..	tresillo-b	x.....x...x.....
gahu	x..x..x...x...x.	tuareg	x...x.x.x...x...
ghana-clap	x.x.x.x..x.x.x..	venda-a	x..x.x..x.x.
guajira-clap	x..x..x.x.x.	venda-b	x.xx.x...x..
ibibio	x.x.x..x.x.x	venda-clap-a	x.xx.x..x.x.
joel-haak	xxxx.x.xx.x.	venda-clap-b	x.x.x.xx.x..
kassa	xx.x.xx.x.x.x.x.	yoruba	x.x.xx.x.xx.

Table 4.5: *Indian* Decitalas.

Name	Rhythm	Name	Rhythm
abhanga	x.x.....	mantha1	x.x.x...x.x.x.x.
abhinanda	x.x.xxx...	mantha2	x...x.x.x.....x.x.
ananga	x.x.....x.x.x...	mudrita3	x...x.x.x.x.x.x.
bindumali	x...xxxxx...	mudrita4	x.x.x.x.x...x.x.
carabhalila	x.x.xxxxx.x.	mukunda	x.xxxxx...
carngadeva	xxx..x....x..x..x.	nandana	x.xxx...
caturasravarna	x...x.xxx...	nandi	x.xxx.x.x...x...
caturmukha	x.x...x.x.....	niccanka	x.x..x..x....x..x..x..x.
crikirti	x.x.x...x...	nihsaruka	x.....x.....
crinandana	x...x.x.x.....	pratimanthaka	x.x.x...x...x.x.
criranga	x.x.x...x.x.....	pratitala	x.xx
darpana	xxx...	pratyanga	x...x...x...x.x.
dhatta	x.x.xxx.x...	rajacudamani	xxx.x.x.xxx.x...
dhenki	x...x.x...	rajamriganka	xxx.x...
dipaka	xxx.x.x...x...	rajanarayana	xxx.x...x.x...
dombuli	x..x..	rajatala	x...x.....xxx...x.x.....
dvitiya	xxx.	rajavidyadhara	x.x...xx
gaja	x.x.x.x.	ranga	xxxxx...
gajalila	x.x.x.x..	rangabharana	x...x...x.x.x.....
gauri	x.x.x.x.x.	rangapradipaka	x...x...x.x...x.....
hamsalila	x..x..	rangodyota	x...x...x...x.x.....
hamsanada	x.x.....xxx.....	rati	x.x...
jaya	x.x...x.x.xxx.....	ratilila	x.x.x...x...
jayacri	x...x.x...x.x...	rayavankola	x...x.x...xx
jayamangala	x.x.x...x.x.x...	sarasa	x.xxxx.x.
kaladhvani	x.x.x...x.x.....	sarasvatikanthabharana	x...x...x.x.xx
kandarpa	xxx.x...x...	shattala	xxxxxx
kanduka	x.x.x.x.x...	simha	x.xx.x.x.
kankalakhanda	xxx...x...	simhanada	x.x...x...x.x...
kankalapurna	xxxxx...x.	skanda	x...x.x...xxx...x...
kankalasama	x...x...x.	tribhangi	x.x.x...x...
kankalavishama	x.x...x...	tribhinna	x.x...x.....
karanayati	xxxx	tryasravarna	x.x.xxx.x.
kirti	x.x.....x...x.x.....	udikshana	x.x.x...
kokilapriya	x...x.x.....	utsava	x.x.....
kollaka	x.x.x...x...x.x.	vanamali	xxxxx.xxx...
kudukka	xxx.x.	vardhana	xxx.x.....
kumuda1	x.xxx.x.x...	varnabhinna	xxx.x...
kumuda2	x.xxxxx...	varnamanthika	x.x.xxx.xx
kuvindaka	x.xxx...x.....	varnayati	x.x.xx
lalita	xxx.x...	vasanta	x.x.x.x...x...x...
lalitapriya	x.x.x...x.x...	vijaya	x.....x...x.....
lila	xx.x.....	vijayananda	x.x.x...x...x...
madana	xxx...	vilokita	x...xxx.....
makaranda	xxx.x.x.	viravikrama	x.xxx...
mallikamoua	x.x.xxxx		
candrakala	x...x...x...x.....x.....x.....x.		
janaka	x.x.x.x.x...x...x...x.x.x...		
dvandva	x.x.x...x...x...x.x.....		
parvatilocana	x...x...x...x.x.....x...x...xx		
simhavikrama	x...x...x...x.x.....x...x.....		
simhavikridita	x.x.....x...x.....x.....x...x.x...x.....x.x.....		

4.2.3 North Indian *Talas*

The final cultural set of rhythms is also from India. However, these are from North India, called the North Indian *Talas*. These rhythms are from Martin Clayton [20], who describes them as the most popular rhythms in North India. Thus we use all the rhythms indicated by Clayton, except those which have a prime number of pulses (discarding only two). Also, we use the clapping patterns to for the North Indian *Talas* in Table 4.6.

Table 4.6: *North Indian* Talas.

Name	Rhythm
ada cautal	x.x...x...x...
adi tal	x...x.....x...
brahma tal	x.xx.xxx.xxxx.
cautal	x...x...x.x.
dhamar tal	x....x....x...
dipcandi tal	x.x.x.....x...
ektal	x...x...x.x.
jhaptal	x.x....x..
jhumra tal	x..x.....x...
matta tal	x.xx.xxx.
pancam savari tal 1	x...x.....x..
pancam savari tal 2	x...x.....xx.
sultal	x...x.x...
tintal	x...x.....x...

4.3 Random

The rhythms presented in this section, have not been used in psychological studies and are not taken from a specific cultural region. The rhythms here are arbitrary in nature as they've been generated at random. Thus, a data set of random rhythms with 12 pulses and another set of random rhythms with 16 pulses are presented below.

The process by which the data sets have been generated can be thought of by a well-known example in probability: flipping a fair coin. Consider generating a rhythm with n pulses. To do so, we simply flip a coin n times, sequentially writing down the result of each coin flip, whether it was heads or tails. After we flip the coin n times we look back at our list of outcomes. Then we make a new list, which will be our random rhythm. Thus, for each outcome of the coin flip, we mark an onset in the new list when the flip resulted in heads, and we mark a silence in the new list when the flip resulted in tails. This is done n times, and when we have finished, a random rhythm of onsets and silences has been created. Now we can repeat this process for as many rhythms as we would like to create and also for any number of pulses n.

However, here we generate rhythms this way first fixing the number of pulses at 12. Thus we have a list of 50 random rhythms with 12 pulses listed in Table 4.7(a). Also, we

fix the number of pulses at 16 and repeat the random rhythm generation process to create another list of 50 rhythms. This list is presented in Table 4.7(b) below.

Table 4.7: *Random rhythm data sets of 12 pulses in (a) and 16 pulses in (b).*

(a) 12 pulse rhythms, 6.02 average onsets.

No.	Rhythm	No.	Rhythm
1	.x..x.xxx..x	26	.xx.xxxxxxx.
2	.xxx.x.x.x..	27	..x.x....xx.
3	xxx.x.xx.x..	28	x..xx.x.x..x
4	xx.x.xx.....	29	...x.xx....x
5	xx.xx.x.xx..	30	.x.xxx....x.
6	xxxxx..x..xx	31	..x.x....x..
7	xx.xx...xxx.	32	.xxxx.x.xx..
8	x..x.x...xxx	33	xx.xx.xxxxx.
9	..x.x.xx.x.x	34	x.xx..xxxxxx
10	x.....x.x.xx	35	..x..x.....x
11	..xx...x..xx	36	..x.xxxxx.x.
12	.xx.x...x...	37	xx..x.xx..x.
13	xxxxxxx...xx	38	..x.xx..xx..
14	.x.xxxxx....	39	..x....x....
15	xx.x.x.xxxx.	40	..xxx....xx.
16	...x.xxxx.xx	41	.x..x.xxx.xx
17	.xxx.x..x.xx	42	.x..xx...x.x
18	x.........x	43	.....xx.xxx.
19	x.xxx..xxx..	44	....x.xx....
20	.xxxx.x.x...	45	xxxxx...xxxx
21	x.xxxx...x..	46	..x.x..xxx..
22	x...x.x.xxx.	47	xx.x..xx..x.
23	.xxx....xx..	48	.x.xx..xxxxx
24	x..x..xxx.x.	49	.x.xxxxx.xx.
25	.xx.xxx.xx.x	50	.x..x..xx.xx

(b) 16 pulse rhythms, 7.94 average onsets.

No.	Rhythm	No.	Rhythm
1	xxxx..xx.xxxxx.x	26	x.........xxx...
2	.xx.xxxx..x.xxx.	27	...xxx.x..xxx...
3	x...x.x..xx.x.xx	28	xx..x..x...xx.x.
4	x..xx.x...x.xx.x	29	x.x.x.x..x...xx.
5	.x.......xxx.x.x	30	xxxxx...xx.x....
6	xxxxx.x.xxxx.xx.	31	x...xx.x..x...x.
7	xxx.xxx.xx..xxx.	32	.xxxx.x...x.xxxx
8	..x.xx..xx.....	33	.....xx.x.xx..xx
9	xxx.x.xxx.xx....	34	x.xxxxx...x.....
10	.....xxx.xx....x	35	xx..x.x.x.x..xx.
11	xx.xxx.xx.x.....	36	.xx.x.x.xx..xx..
12	...xx.xxx.xx..x.	37	.xx..xxxxx....xx
13	x..x.x.x.x...x.x	38	xxx..x.x...x...x
14	.x..xxx.xx.xxxx.	39	x....x.x..x..xxx
15	.xx.x........xxx	40	x..xx.....x.xx..
16	xxxxx....xx..xx.	41	....xx.xx..x.xxx
17	x..x....x.xx.xxx	42	.xxxxx.xxx..x.x.
18	.xx..xx...x.xx.x	43	xx.xxxx..xxx.xx.
19	x..xxx.......x.x	44	xx....xxx.x.xx..
20	xxxx.x.x.xxx..x.	45	....xx...xxxx.xx
21	..xxx....xx..x.x	46	.xxx....xx.x.xxx
22	xx...x.x.x.xxxxx	47	....x..x.xx.xx..
23	.xx.xx....x.x.xx	48	x.x.x....xx.xx..
24	x.xxx.x..x....x.	49	.xxx.x.....x...
25	x..xxxx.x.x....x	50	.xxx..x..xxxx.xx

Granted that generating 100 rhythms by this coin-flipping process would prove to be a bit tedious, thus the rhythms used in this thesis and shown in Tables 4.7(a) and 4.7(b), have been generated in this manner described by a computer program which outputs a 0 or 1 at random, where the probability that a 0 is output is $1/2$ and the probability that a 1 is output is also $1/2$; i.e., using a uniform distribution. There are many uniform random number generators to choose from; however, the authors of *Numerical Recipes in C* [123], point to Pierre L'Ecuyer's work [93, 94, 95]. Thus, an implementation by L'Ecuyer hosted (and also recommended) by Luc Devroye [39, 92], was used.

Chapter 5

Methodology

The methodology is now described, which uses the complexity measures and experimental data previously presented. The first section describes the evaluation process of the complexity measures regarding the human-based measures of rhythm complexity, and the variety of data sets already seen. The second section discusses the visualization of the evaluation process for analysis purposes. The third section then presents an implementation developed to incorporate the evaluation technique and prime the results for the visualization method.

5.1 Evaluation

Since one of the goals in this thesis is to determine the overall performance of the rhythm complexity measures, a suitable method for evaluation must be devised. In order to attain the overall performance, there are two aspects in the evaluation which must be considered. First, each complexity measure is validated against psychological experimentation, which takes human-based measures of rhythm complexity as a baseline. Second, all pairs of complexity measures are compared to determine if the measures attain similar or different results. In other words, we want to gauge the relationship among all pairs of rhythm complexity measures, and in some cases, this includes human-based measures of rhythm complexity.

Taking into consideration that there is a mix between empirical (i.e., human-based measurements) and computational rhythm complexity measures, a general evaluation method is followed. Thus, this thesis follows the method of *correlation analysis* [24] to determine the relationship between the complexity measures. By using such a general method, which is common among the behavioral sciences (e.g. psychology), we may adequately evaluate the wide-range of rhythm complexity measures, which include psychological and computational approaches.

5.1.1 Correlation Analysis

As stated by John Carroll in 1961, the correlational method is used "as a way of measuring something called 'relationship' between variables" [15]. Indeed, this notion of a relationship between variables can be explored by statistical correlation [75, 87, 152]. In this work, each rhythm complexity measure is considered such a variable, whose relationship shall be explored. In fact, Chen and Popovich state that "[to] evaluate if one assessment tool ... is superior to others" [19], is appropriate practice for using correlation [19]. Here, we want to determine which measure of rhythm complexity performs best regarding human-based measures of complexity. Thus using correlation to discover such performance

is ideal. In addition, calculating the correlation between all pairs of complexity measures provides our second evaluation criterion of evaluating whether measures perform similarly or not.

The literature on correlation reveals the *Pearson product-moment correlation coefficient* [116, 117] to be the most widely used [19] and well-known [19] method for measuring the correlation between two sets. Consider Equation (5.1), which gives the formula for computing the Pearson product-moment coefficient (denoted by r) for sets X and Y. Let there be n total elements in each set, and let $\overline{X}$ to be the mean of the values in X and $\overline{Y}$ to be the mean of the values in Y.

$$r = \frac{\sum_{i=0}^{n-1}(x_i - \overline{X})(y_i - \overline{Y})}{\sqrt{\sum_{i=0}^{n-1}(x_i - \overline{X})^2 \sum_{i=0}^{n-1}(y_i - \overline{Y})^2}} \tag{5.1}$$

However, before choosing the Pearson product-moment correlation coefficient, let us explore whether such a method will adequately meet our goals. First we must consider the nature of our variables, and whether the values are suitable for calculating Pearson's r. For instance, if the values were continuous, then Pearson is indeed the obvious choice; however, if there are variables containing ordinal values, then this choice is not so clear [19].

In our case, there are variables which contain ordinal values. The reason for this is because the precise human measurements were not available (e.g. Povel and Essens [119]) and so in order to avoid discrepancies, ordinal values were obtained. Based on this alone, it seems unreasonable to apply the method of calculating the Pearson product-moment correlation coefficient. Moreover, as pointed out by Charles Spearman in 1904, this coefficient assumes that a linear relationship exists between the variables [152]. This linear relationship may or may not exist; however, since we are using complexity measures for our variables which present a wide variety of calculations, generality in our correlation coefficient is preferred. To attain such generality, a special case of correlation, termed *rank correlation* [60, 61, 75, 87, 88, 152], shall be used.

5.1.2 Rank Correlation Analysis

Rank correlation is a method for comparing how different processes arrange objects in an order [87]. More formally, let P be a set of objects. A binary relation $\leq$ placed on P is referred to as an *order* such that the following three properties hold: (1) *reflexivity*, $x \leq x$ for all $x \in P$, (2) *antisymmetry*, $x \leq y$ and $y \leq x$ implies $x = y$ for all $x, y \in P$, (3) *transitivity*, $x \leq y$ and $y \leq z$ implies $x \leq z$ for all $x, y, z \in P$. In this thesis an in-depth discussion regarding order theory is not necessary; however, an insightful book by Davey and Priestley [36] is recommended for those who are curious. The previous definition presented was from that text.

To put the definition of order into context, let us define the previously mentioned term *process*, to be a systematic series of actions used to quantify some property of an object. Hence, the processes here are the algorithmic measures presented in Chapter 3, the property being measured is complexity, and the objects are musical rhythms. Since the algorithmic measures yield real values representing the complexity of a musical rhythm, when a measure is applied to a set of rhythms, a set of real complexity values is produced. This naturally admits an order. Therefore, we may use rank correlation coefficients to quantify the relationship between how two such subsets are ordered.

The interpretation of what such a rank correlation coefficient means can have different interpretations. For instance, in our case, we artificially convert all the scores to ranks, so that all the rhythm complexity measures can be equally compared. This avoids the need for scaling or developing a method for comparing ordinal to continuous values; as would be the case if rank correlation had not be chosen. In addition, the advantages of converting all scores to ranks can be seen when outliers are present. When data sets are small, outliers can potentially have a large impact, and since the rhythm data sets we are working with can potentially be labeled as *small*, the outlier problem is important. By converting to ranks, outliers are avoided [19]. Moreover, with artificial conversion to ranks, a monotonic relationship becomes apparent between the random variables, thus they can be described more generally (positively or negatively) in a non-linear way. However, a main disadvantage to using purely ranked data is that the quality of ranking is limited [87]. This is because no information about the distance between adjacent objects in the ranking is known. Although, in a sense, this *disadvantage* can be seen as an advantage relating to the problem of outliers previously discussed.

The advantages to using rank correlation analysis over Correlation Analysis are clear; however, the literature [19, 75, 87, 88, 91, 117, 152, 153], presents two main choices for rank correlation coefficients: the *Spearman rank correlation coefficient* and the *Kendall tau rank correlation coefficient*. At the most basic level, the Spearman rank correlation coefficient uses the sum of squared differences in the rank values between the two variables, and the Kendall tau rank correlation coefficient assesses the agreement (concordant pairs) between the ranks [87, 88, 111]. Choosing between the coefficients is not clear in the literature. Some argue that Spearman's coefficient has the advantage by being a direct derivation of the Pearson's r, thus retaining the interpretability of Pearson's coefficient. Others argue that Kendall's tau retains the advantage because the proportion of the concordant and discordant pairs in the method lends a more direct interpretation [111]. This thesis uses the Spearman rank correlation coefficient as the method for rank correlation analysis because of its natural derivation from the Pearson product-moment correlation coefficient, and because

it is the most widely used coefficient [19].

The most common formula for calculating the Spearman rank correlation coefficient, denoted by r_s, is shown in Equation (5.2) [75]. Also, n denotes the number of values, and $\mathbf{d}$ contains the differences between the rankings of corresponding values from one variable compared to another, or in our case, from one complexity measure compared to another.

$$r_s = 1 - \frac{6 \sum_{i=0}^{n-1} d_i^2}{n^3 - n} \tag{5.2}$$

However, there is a case that this equation, as originally formulated, does not accurately handle. Namely, the Spearman rank correlation coefficient must be adjusted in the case that the ranking has ties, i.e., two or more items have the same rank. Adjusting for ties is common [87] among the rank correlation coefficients, and shall be discussed in § 5.3. Next we describe a method for visualizing the correlation given by the Spearman rank coefficients. Since we have a large number of complexity measures, we compute r_s between *all pairs* of the complexity rankings of the rhythms according to the measures. For example, if we have complexity values from Toussaint's Metrical Complexity measure, the Longuet-Higgins and Lee Complexity measure, and Pressing's Cognitive Complexity measure, on some list of rhythms, we calculate r_s between the all pairs of complexity rankings from these measures. This yields a symmetric square matrix of coefficients we denote as $\mathbf{R_s}$, which tells us the relationship between any pair of complexity rankings from two measures. Consider how such relationships are visualized below.

5.2 Visualization

In the previous section, we calculated Spearman rank correlation coefficients between all pairs of rhythm complexity measures and composed a square symmetric matrix. This matrix reveals the relationships between the complexity measures; however, such a relationship is perhaps not clearly depicted. To illustrate the relationships, the coefficient matrix is first converted into a distance matrix and then this distance matrix is visualized as a phylogenetic tree, as used in the field of bioinformatics for phylogenetic analysis.

5.2.1 Distance Matrix

A Spearman rank correlation coefficient may be considered as a distance by applying the following formula. Below, consider $\mathbf{R_s}$ to be a symmetric matrix of coefficients and the matrix $\mathbf{D_s}$ to be the corresponding distances.

$$\mathbf{D_s} = 1 - \mathbf{R_s} \tag{5.3}$$

Equation (5.3) is a typical approach which has been used by clustering software and the open-source C Cluster Library [38]. Moreover, this approach is supported more generally

where $\mathbf{R_s}$ may be subtracted from any constant [21]. In addition, another approach, which is mentioned by Everitt et al. [49], is to divide Equation (5.3) by two in order to normalize to an interval between zero and one (inclusive). However, we choose to avoid scaling, and hence follow the approach of de Hoon, et al. [38] and Cliff et al. [21] (see Equation (5.3)) to convert the matrix of Spearman rank correlation coefficients to their corresponding distance values. Moreover, $\mathbf{R_s}$ lends itself to comparing distances because of its invariance under monotone transformations [41]. Upon the conversion to $\mathbf{D_s}$, the range of the distances lies in the inclusive interval $[0, 2]$. We arrive at a distance of zero when there is perfect rank correlation (a positive one), and we arrive at a positive two when there is perfect inverse rank correlation (a negative one). Converting the Spearman rank correlation coefficients to a distance is sufficient [3, 118] for a visualization method, which is typically applied to distance values; i.e. *cluster analysis*.

5.2.2 Cluster Analysis

Up to this point we have Spearman rank correlation coefficients which have been converted to distances, and thus compose the distance matrix $\mathbf{D_s}$. Moreover, recall that these distances represent dissimilarity between rhythm complexity measures; i.e., the closer the distance value is to zero, the more similar (more correlated) the performance of the complexity measures. Given this, we want to present the information inherent to the distance matrix, depicting the relationship between all the complexity measures. Cluster Analysis can be used to convey the relationship between complexity measures by revealing the underlying structure of the data [44]. Thus, we first consider different clustering approaches and then evaluate which method most suits the data under consideration.

Cluster analysis is a widely used technique found in many different fields; e.g., in biology as *numerical taxonomy*, psychology as *Q analysis*, and artificial intelligence as *unsupervised pattern recognition* [49]. Hence, there is an array of methods which can be applied to analyze data by way of clustering. However, the goal of cluster analysis remains the same:

> "[To explore] data sets to assess whether or not they can be summarized meaningfully in terms of relatively small number of groups or clusters of objects which resemble each other and which are different in some respects from the objects in other clusters" [49].

In order to determine a suitable method for cluster analysis, perhaps we should turn to our data. Given our distance matrix $\mathbf{D_s}$ of dissimilarities, we would like a method which fits points into a two-dimensional space where the distance between the points match (as closely as possible) to the dissimilarities in $\mathbf{D_s}$. This, in a sense, can generally be termed as *multidimensional scaling*; to embed variables as points in a coordinate space such that

the distances between the points represent the observed proximity [49]. A common cluster analysis technique which can provides a two-dimensional planar embedding from a distance matrix is *hierarchical clustering* [44]. Specifically, the use of *dendrograms* provides a way to visualize corresponding distances [49]. However, other techniques such as *relative neighborhood graphs* [80] may also be used as a graphical way to cluster data on a planar surface [96, 97].

As Everitt et al. point out, "no particular clustering method can be recommended" [49], thus choosing requires a deeper investigation of the data at hand to determine the most suitable approach. From above, we know that because of our choice of formulating Spearman rank correlation coefficients as distances, methods such as hierarchical clustering can appropriately be used. In addition to this, the nature of our data is culturally based; rhythm is undoubtably cultural. As Ian Cross states, "...a clean dissociation between culture and biology–or between music and evolution–is unfeasible" [35]. Because of the proposed intrinsic connection among music, culture, evolution, and biology, perhaps the most appropriate method for cluster analysis may be found in the biological domain.

One method stemming from bioinformatics and computational biology is *phylogenetic analysis* [16, 43, 65, 72, 73, 78, 79], which has been applied to rhythms [40, 68, 164, 165, 166, 168, 169], as well as other cultural phenomena (e.g. textile patterns [102, 113, 160]). In fact, phylogenetic analysis is a generalization of the dendrogram technique in hierarchical clustering [49]. Essentially, it is an *additive tree* [49] where the lengths of the edges between notes represent the distances in a dissimilarity matrix, such as our $\mathbf{D_s}$. Here we analyze such trees produced from phylogenetic analysis, and thus we term this approach to be *phylogenetic tree analysis*.

5.2.3 Phylogenetic Tree Analysis

In bioinformatics and computational biology, a *phylogenetic tree*, "depicts the evolutionary relationships of a target set of sequences" [72]. In this thesis, we shall be depicting the relationships between musical rhythm complexity measures as determined by calculating Spearman rank correlation coefficient distances using sequences of musical rhythms. The construction of such a phylogenetic tree generally involves a four-step process: (1) selecting of the variables to be compared (musical rhythm complexity measures), (2) measuring the relationship (in terms of rhythm complexity by way of Rank Correlation) of the variables, (3) creating a distance matrix based on the measurement of relationship, (4) applying a clustering technique to generate the tree [72].

The clustering techniques available for step four vary in nature; however, they all maintain a minimalist ideal. That is, they strive for the simplist pathway from one node in the tree to the next [72]. The earliest method for phylogenetic tree construction known as the

Parsimony method was proposed by Edwards and Cavalli-Sforza [16, 45, 46] and implemented by Camin and Sokal [14]. The process behind the Parsimony method is to begin by placing the internal-most nodes in the tree and then working outwards, placing nodes which have the minimum distance compared to the internal nodes [14]. By doing so, each node's placement is examined and thus when the number of variables is large, the number of topologies grows quite quickly. Moreover, the Parsimony method is rather specific to those variables which are closely related diverging from a common source [72]. In the case of biology, this means biological sequences, like DNA, which evolve from the same lineage. Thus, when there is large evolutionary change in different lineages, the Parsimony method converges to the wrong phylogenetic tree [51]. Since we do not consider lineages or the meaning of this for musical rhythm, we prefer a more general method to constructing a phylogenetic tree.

Robust alternatives include the *Maximum-Likelihood* method [50], *Fitch-Margoliash* method [54], and *Neighbor Joining* (NJ) method [137]. The Maximum-Likelihood method calculates the probability of obtaining the data under consideration, given a phylogenetic tree of a certain topology using information from a general evolutionary model [50]. Thus, the topological space of trees is searched to maximize the likelihood for obtaining the given data. The Maximum-Likelihood method can be computationally expensive with large data sets, and may fall into a local maxima, therefore not arriving at the true maximized likelihood [72]. This method also does not seem appropriate for this thesis because of the potentially expensive computational costs and also because we do not develop an evolutionary model for musical rhythms.

The Fitch-Margoliash method is a more general approach, which is a bottom-up technique for creating a phylogenetic tree [54]. The data is first clustered into small subsets, then each subset is joined based on the two subsets which admit a minimum distance, found by taking the average among pairs of elements between each subset [54]. Thus, the Fitch-Margoliash method follows the *Unweighted Pair-Group Method with Arithmetic mean (UPGMA)* [49], to join clusters. In addition, this method strictly uses the distance matrix and does not rely on evolutionary information. However, in the words of Fitch and Margoliash, "because average mutation distances are now being used, the solutions obtained [in terms of phylogenetic trees] are very unlikely to permit an exact reconstruction of the input data" [54]. Therefore, even though this method fits our requirement of purely using the distance matrix for phylogenetic tree construction, perhaps there is a more robust method to handle all distance-types. In fact, the NJ method proposed by Saitou and Nei is just that, and is praised in the Bioinformatics/Computational Biology community [72].

The Neighbor Joining (NJ) method is a robust way to construct a phylogenetic tree, by

joining pairs of nodes which minimize the total branch length of the entire tree at each stage in the clustering [137]. Each variable is represented by a node in the tree, and is termed a *neighbor* if they are joined by an internal node in the tree. This process of joining neighbors is continued until the tree is generated. At each step in the process, nodes are joined in a way to minimize the total branch length of the tree [137]. The NJ method produces a correct, purely additive tree, and also out-performs the other method discussed above [137]. However, in 1997, Olivier Gascuel proposed an improved version of the NJ method called the BIONJ method [65]. The main advantage of BIONJ is that the algorithm does not make the assumption that the total branch length of the tree should be minimized. By making this assumption, the NJ method produces an under-estimate of the real phylogenetic tree [65]. Since the BIONJ method avoids this assumption, the algorithm outperforms the NJ method. Moreover, BIONJ still retains all of the benefits of the NJ method, such as high reliability and low time complexity of $O(n^3)$, c.f. $O(c^n)$ of the Parsimony method [72, 83]. Therefore, we use the BIONJ method for constructing the phylogenetic trees in this thesis.

5.3 Implementation

The details of the implementation for our method is discussed. The implementation of the rhythm complexity measures is first described. Then the graphical user interface developed for computing the Spearman rank correlation coefficients and significance matrices is described. Finally, the program used to visualize such matrices is discussed.

5.3.1 Rhythm Complexity Measures

Each of the rhythm complexity measures were programmed in the C programming language [89] on the Gentoo Linux [66] platform, complied with GCC v4.1.2 [59]. The C programming language was chosen for performance and also because this language is widely accepted and practiced in the programming community. Each complexity measure has been individually coded for modularity to be used by the graphical user interface.

5.3.2 Python GUI

The Python programming language [127] and the Python graphics toolkit [162] were used to develop a GUI (graphical user interface) to act as a wrapper for the rhythm complexity measures. This provided a convenient way for calculating Spearman rank correlation coefficients. Figure 5.1 illustrates the GUI.

The interface consists of three main steps. First, musical rhythm data (in box notation) must be imported to the application's data table, either one at a time or by loading a *comma separated value* (CSV) file. Figure 5.1 shows 15 rhythms, which have already been loaded. Additionally, as seen in the figure, columns may be present in the CSV file containing human-based rankings. Second, one must compute the complexity of the rhythms using the rhythm complexity measures. The user may select measures to include in the computation.

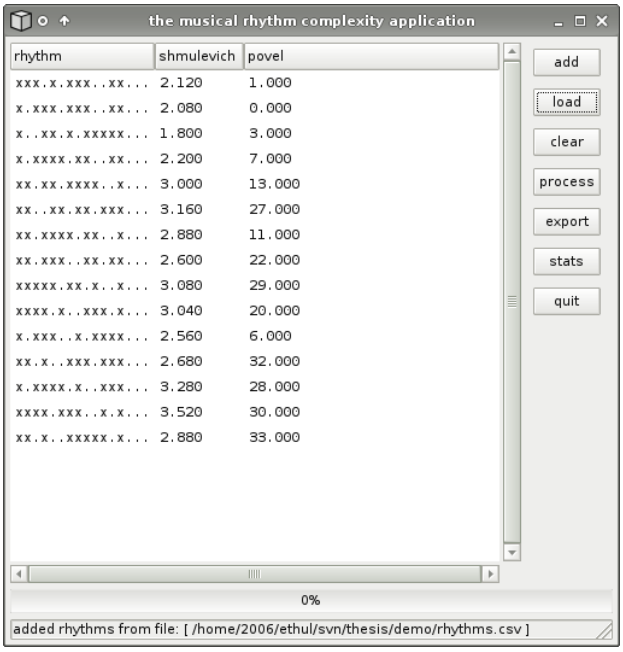


Figure 5.1: *GUI for measuring rhythm complexity.*

Each selected measure will appear as a columns next to the rhythms. Thus, one may see for any measure the complexity of a give rhythm. Third, statistical analysis is performed. This process is detailed below.

Spearman Rank Correlation

Spearman rank correlation coefficients are computed between all pairs of rankings of the rhythms of the selected measures. The calculation of r_s is modified from the version presented before. Here we adjust for tied rankings (i.e., two rhythms have the same complexity according to one measure), and we account for large sample approximation [74]. The modified version for calculating the Spearman rank correlation coefficient follows from Hollander and Wolfe [74]. Equation (5.4) shows the formula denoted by r_s^*.

$$r_s^* = \frac{n(n^2-1) - 6\sum_{s=0}^{n-1} d_s^2 - \frac{1}{2}\left\{\sum_{i=0}^{g-1}[t_i(t_i^2-1)] + \sum_{j=0}^{h-1}[u_j(u_j^2-1)]\right\}}{\left\{[n(n^2-1) - \sum_{i=0}^{g-1} t_i(t_i^2-1)][n(n^2-1) - \sum_{j=0}^{h-1} u_j(u_j^2-1)]\right\}^{1/2}} \tag{5.4}$$

In the equation, n is the number of values in ranked lists and g is the number of tied groups, where vector $\mathbf{t}$ contains the number of elements in each tied group for the first ranked list. Similarly, h and vector $\mathbf{u}$ represent the information regarding the tied groupings for the second ranked list. We should note that the adjustment for ties produces an average rank value for all those ranks in a tie group [74].

Statistical Significance

In addition to the Spearman rank correlation coefficients, corresponding significance of each coefficient is computed. Here we test the null hypothesis that the population Spearman rank correlation coefficient, denoted ρ_{rs}, is zero. This significance is an approximation of the Student's t-distribution [19]. This is a stable approximation because it does not depend on the original distributions of the variables [123]. Thus we calculate a two-sided significance level, which tests the null hypothesis for positive and negative coefficients. The code used for this calculation is from Gary Strangman [155], who is a member of the Neural Systems Group at Harvard Medical School. Note that he derived his Python implementation from *Numerical Recipes in C* by Press et al. [123]. The significance calculated shows the confidence level, where a higher value means higher confidence that the null hypothesis may be rejected. The p-values may be found by subtracting the confidence level from 1.

5.3.3 SplitsTree

The Python GUI generates a distance matrix from the Spearman rank correlation coefficients to be used for phylogenetic tree analysis. The BIONJ method is used for the construction of such phylogenetic trees. Since this is a well-known and popular method, there are many existing programs [52] which may be used. The program we chose for phylogenetic Tree generation is *SplitsTree* [78, 79]. SplitsTree was developed by Daniel Huson who is a Professor of Algorithms in Bioinformatics at the University of Tübingen, and David Bryant who is an Associate Professor of Mathematical Biology at the University of Auckland. In addition, Bryant was previously a Professor at McGill University. We support their software as it is one of the leading tools in computing evolutionary networks [178].

We use *SplitsTree4* version $4.9.1$, released April 15, 2008, which we refer to as *SplitsTree* in this thesis. SplitsTree implements the BIONJ method for phylogenetic tree construction according to the algorithm proposed by Gascuel [65]. In order to quantify the accuracy of a phylogenetic tree construction, the method of *least-squares fit* [16, 43, 156, 184] is used within SplitsTree. The method of least-squares fit was proposed by Tanaka and Huba [156], where we use the definition by Winkworth et al. [184] as follows in Equation (5.5). Let $\mathbf{D}$ be the distance matrix and $\mathbf{P}$ be a matrix of edge distances between all terminal nodes in the phylogenetic tree generated; we sum over all pairs i, j.

$$\text{Least-Squares Fit} = \left[1 - \frac{\sum_{i,j}(P_{i,j} - D_{i,j})^2}{\sum_{i,j} D_{i,j}^2}\right] * 100 \tag{5.5}$$

Thus, the methodology of this thesis is implemented by combining the rhythm complexity measures, the Python GUI, and SplitsTree. These tools are used to generate the results to be discussed.

Chapter 6

Results

The results presented consist of three main parts. The first part describes phylogenetic trees regarding the psychological data of *human performance complexity*, *human perceptual complexity*, and *human metrical complexity*. These human-based complexity measures are derived from psychological experimentation of Povel and Essens [119], Essens [48], and Fitch and Rosenfeld [57]. The second part of the results compares the rhythm complexity measures to rhythmic patterns of African and Afro-Cuban *Timelines* which contain either 12 or 16 pulses. Thus, we are testing whether the complexity measures are robust to rhythms with a different number of pulses. Indian *Decitalas* and North Indian *Talas* are also used for testing the measures of a varying number of pulses. The third part presents the results of data which are random. Stemming from the previous part of the results, the robustness of the measures are tested on randomly generated rhythms with different number of pulses. The goal is to validate the cultural robustness in a more general sense.

6.1 Psychological

The results discussed below pertain to the human-based performance complexity, perceptual complexity, and metrical complexity. These measures have be acquired from four psychological studies, as previously described.

6.1.1 Povel and Essens; Shmulevich and Povel

Results using the data set from Povel and Essens [119], permit validation of the complexity measures against two types of human-based complexity measures from psychological experimentation. Povel and Essens studied human rhythm reproduction and determined how difficult it was to reproduce each of the rhythms in their data set. We use this as a human-based performance complexity measure where each rhythm obtains an ordinal rank of complexity, according their study. In addition, the second type of complexity seen here is a human-based measure for perceptual complexity. This is from Shmulevich and Povel's [147] psychological study on human judgments of rhythm complexity. We use this as a measure for human-based perceptual complexity of the rhythms in this data set, where ordinal ranks are obtained from their study. Figure 6.1 shows a BIONJ phylogenetic tree visualizing the Spearman coefficient distance matrix. Nodes labeled *human performance complexity [povel and essens]* and *human perceptual complexity [shmulevich and povel]*, correspond to *Povel and Essens' Human Performance Complexity* and *Shmulevich and Povel's Human Perceptual Complexity*. Each other node corresponds to a complexity measure.

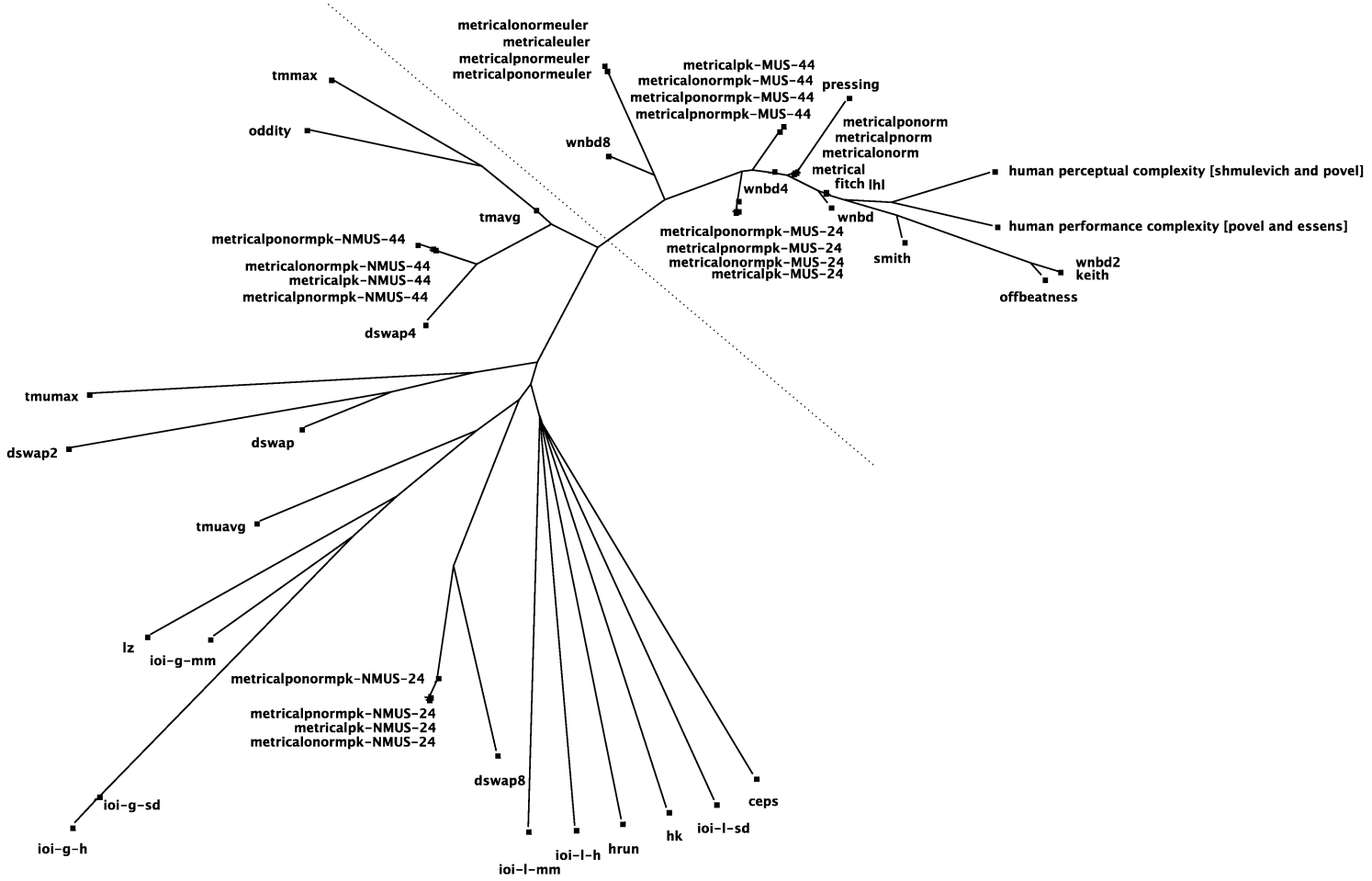


Figure 6.1: *BIONJ phylogenetic tree of the complexity measures, and the Povel and Essens' Human Performance and Shmulevich and Povel's Human Perceptual Complexity, least-squares fit 97.4.*

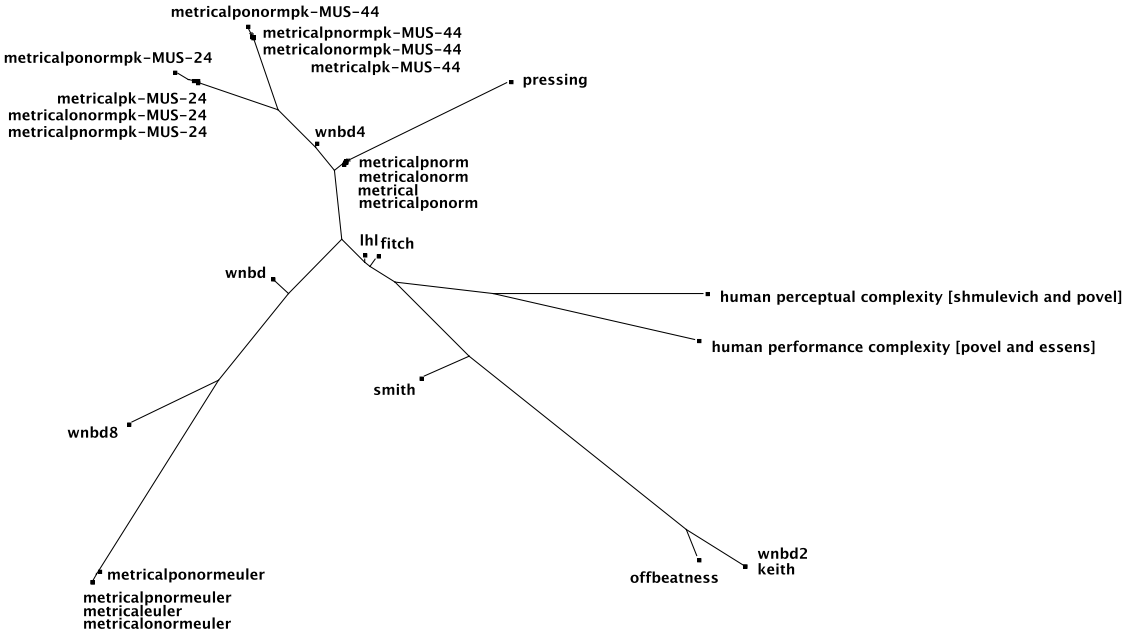
The phylogenetic tree in Figure 6.1 reveals a separation, as shown by the dotted line. Those measures above the line have an average Spearman rank correlation coefficient to the Human Performance and Perceptual complexity measures which is $\bar{r}_s^* \geq 0.500$. The only complexity measure which is placed below the line, but fits this criterion is *tmavg* with $\bar{r}_s^* = 0.534$. Moreover, all measures above the line have significant Spearman rank correlation coefficients for both human measures at the $p < 0.05$ level.

To clearly see how these measures perform with regard to the Human Performance and Perceptual Complexity, consider Figure 6.2(a), which shows those measures above the dotted line in Figure 6.1.

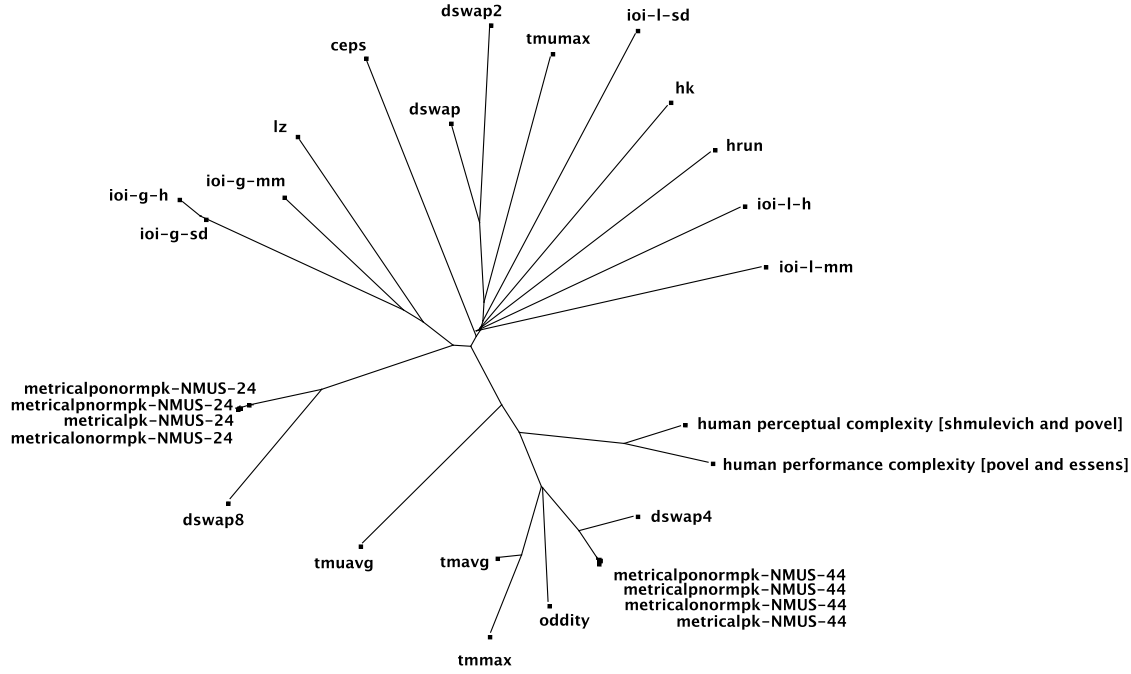
We can see that the measures which are based on the Lerdahl and Jackendoff [99], and Yeston [185] metrical hierarchy are closest to the human judgments. Those measures are: *lhl*, *fitch*, *smith*, and the measures prefixed with *metrical*. Another measure which performs well is the *wnbd* measure and *pressing*. In addition, it is noteworthy that all of the *metrical* complexity measures using the Euler weighting scheme [69] or the weighting scheme of the musicians from Palmer and Krumhansl's [114] study for time signatures $4/4$ and $2/4$, perform closely to the Lerdahl and Jackendoff weighting scheme. As Figure 6.2(a) shows, the *metrical* complexity with the three different metrical weighting schemes is arranged such that the Lerdahl and Jackendoff scheme (*metrical*) is in the middle of the path of the Euler (*metricaleuler*) and Palmer and Krumhansl (*metricalpk-MUS-24* or *metricalpk-MUS-44*) schemes. The correlations between the pairs are: *metrical* to *metricaleuler* is $r_s^* = 0.666$, *metrical* to *metricalpk-MUS-24* is $r_s^* = 0.882$, *metrical* to *metricalpk-MUS-44* is $r_s^* = 0.901$, *metricaleuler* to *metricalpk-MUS-24* is $r_s^* = 0.661$, *metricaleuler* to *metricalpk-MUS-44* is $r_s^* = 0.588$, and *metricalpk-MUS-24* to *metricalpk-MUS-44* is $r_s^* = 0.898$. Moreover, a more general observation is that both weighting schemes of the musicians from Palmer and Krumhansl's study significantly outperform that of the non-musicians.

To provide a better sense of how each complexity measure performed, Table 6.1(a) contains a sorted list of the Spearman rank correlation coefficients of each complexity measure paired with the Human Performance Complexity where $r_s^* \geq 0.500$. Similarly, Table 6.1(b) contains those complexity measures with correlation $r_s^* \geq 0.500$ when paired with the Human Perceptual Complexity. These tables closely reflect the phylogenetic tree in Figure 6.2(a).

Finally, Figure 6.2(b), contains those measure which did not perform well with respect to the human judgments of Human Performance and Perceptual Complexity. However, because the rhythms in this data set each had the same number of onsets and pulses, measures based on entropy are expected to yield poor results. Moreover, since these



(a) BIONJ phylogenetic tree of the complexity measures in the top of the cut, least-squares fit 98.5.



(b) BIONJ phylogenetic tree of the complexity measures in the bottom of the cut, least-squares fit 98.0.

Figure 6.2: *BIONJ phylogenetic trees of the complexity measures in the top and bottom of the cut (see Figure 6.1).*

Table 6.1: *Complexity measures with $r_s^* \geq 0.500$ and significance $p \leq 0.05$ compared to Povel and Essens' Human Performance Complexity and Shmulevich and Povel's Human Perceptual Complexity.*

(a) Performance.

Complexity Measure	r_s^*
smith	0.800
fitch	0.787
lhl	0.787
wnbd	0.747
wnbd4	0.700
metricalpnormpk-MUS-24	0.687
metricalpk-MUS-24	0.683
metricalonormpk-MUS-24	0.683
metrical	0.682
metricalpnorm	0.682
metricalonorm	0.682
metricalponorm	0.682
wnbd2	0.680
keith	0.680
metricalponormpk-MUS-24	0.679
metricalponormpk-MUS-44	0.601
wnbd8	0.600
metricalpnormpk-MUS-44	0.599
metricalonormpk-MUS-44	0.597
metricalpk-MUS-44	0.597
metricalponormeuler	0.595
metricalpnormeuler	0.588
metricaleuler	0.588
metricalonormeuler	0.588
pressing	0.576
offbeatness	0.570
tmavg	0.523

(b) Perceptual.

Complexity Measure	r_s^*
fitch	0.755
lhl	0.755
wnbd	0.738
smith	0.737
metrical	0.694
metricalpnorm	0.694
metricalponorm	0.694
metricalonorm	0.694
wnbd4	0.690
metricalponormpk-MUS-24	0.682
metricalpnormpk-MUS-24	0.677
metricalpk-MUS-24	0.673
metricalonormpk-MUS-24	0.673
wnbd8	0.669
pressing	0.630
metricalpk-MUS-44	0.626
metricalonormpk-MUS-44	0.626
metricalpnormpk-MUS-44	0.625
metricalponormpk-MUS-44	0.605
tmavg	0.545
wnbd2	0.511
keith	0.511
metricalpnormpk-NMUS-44	0.508
metricalpk-NMUS-44	0.506
metricalonormpk-NMUS-44	0.506
metricaleuler	0.503
metricalpnormeuler	0.503
metricalonormeuler	0.503

rhythms were also composed of the same onset intervals, but simply re-arranged, measures based on the interonset intervals also did not perform well. This is why they are shown in a fan shape off the main tree. Additionally, the weighting scheme from Palmer and Krumhansl's study [114] from non-musicians (*nmus*) performed poorly compared to the weighting scheme from the musicians. The average correlation to the two human complexity measures combined with the two meters ($2/4$ and $4/4$) for musicians is $\bar{r}_s^* = 0.645$ and for non-musicians is $\bar{r}_s^* = 0.273$.

6.1.2 Essens

The data set from Essens' study [48], provides two human measures from psychological experimentation. The first measure is termed *Essens' Human Performance Complexity*, and the second measure is *Essens' Human Perceptual Complexity*. Hence, Essens provides a secondary data set for validating the measures of complexity against human-based performance and perceptual complexity.

Consider Figure 6.3 below which depicts a phylogenetic tree for the complexity measures along with both human measures. The distances for generating the tree were derived from the Spearman rank correlation coefficients between all pairs of measures, and the BIONJ algorithm in SplitsTree was used for tree construction.

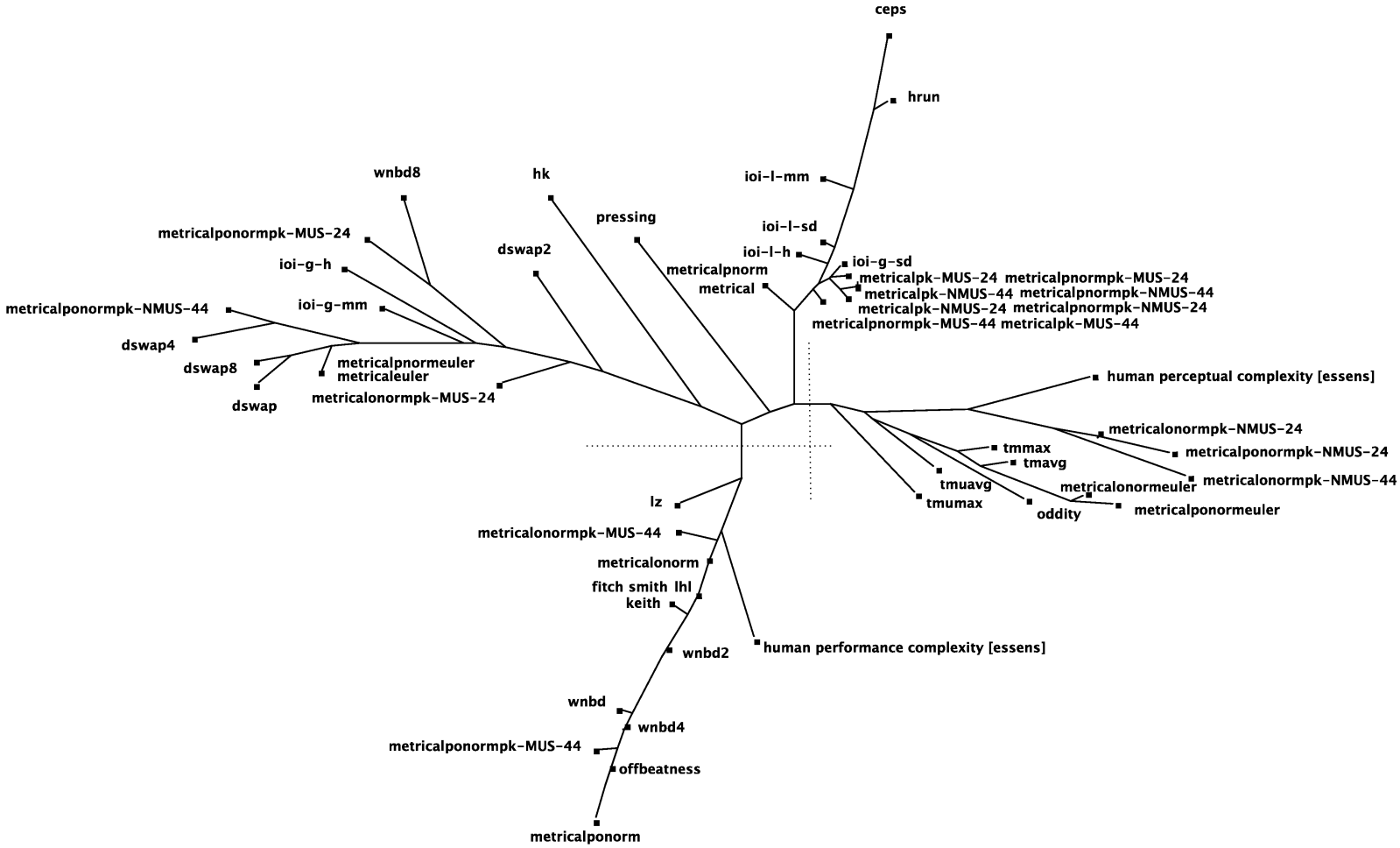


Figure 6.3: *BIONJ phylogenetic tree of the complexity measures, and the Essens' Human Performance and Perceptual Complexity, least-squares Fit 90.0.*

In Figure 6.3, we make two cuts in order to provide a clearer view of the relationships of the top-performing complexity measures to both human-based measures. The first cut is the horizontal line which separates the lower branch of the tree. The second cut is a vertical line, which separates the right-most branch of the tree. Each separated segment contains one of the human complexity measures: Figure 6.4(a) shows those measures closest to Essens' Human Performance Complexity and Figure 6.4(b) shows those measures closest to Essens' Human Perceptual Complexity. Two cuts were chosen because the correlation between the human complexity measures was $r_s^* = 0.015$. Thus, one cut would not suffice.

Regarding Essens' Human Performance Complexity, consider Table 6.2(a), which lists the coefficients greater than or equal to 0.500 and having $p \leq 0.05$. Note that *wnbd* and *metricalponorm* were included by the cut in the tree shown in Figure 6.4(a); however, they have been excluded from the table because their p-value was not at the 0.05 significance level.

The top-performer regarding the human-based performance measure, is Toussaint's Metrical Complexity (Onset Normalized) weighted by Palmer and Krumhansl's musicians-$4/4$ scheme. Also, the top-performers from the previous data set performed worse using this data set. The *lhl, fitch, smith*, and *wnbd*, correlations are just over 0.670, compared to 0.750 using the previous data.

Table 6.2: *Complexity measures with $r_s^* \geq 0.500$ and a significance level of $p \leq 0.05$ compared to Essens' Human Performance Complexity in (a), and since no correlation coefficients were higher than 0.500, when compared to Essens' Human Perceptual Complexity, the measures which correspond to the cut, shown in Figure 6.4(b), are listed in (b), note that* oddity *is the only significant r_s^* with $p \leq 0.05$.*

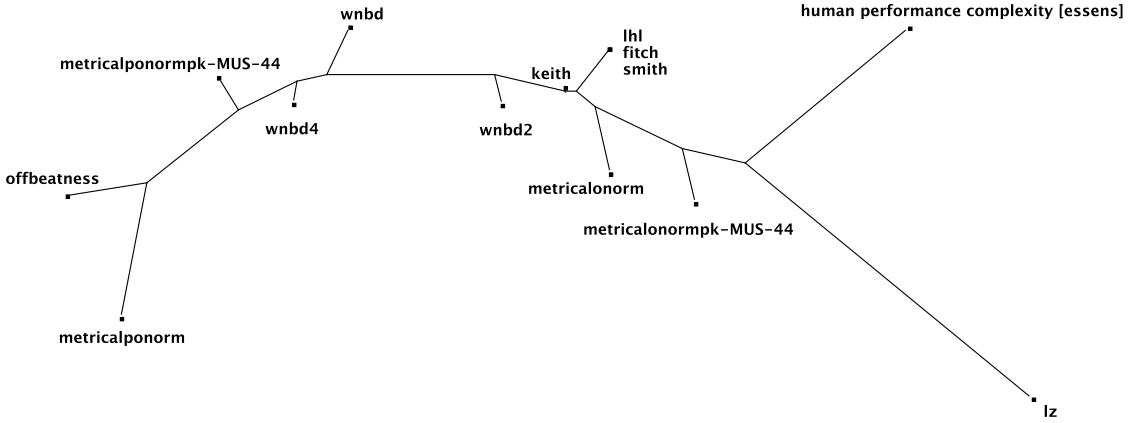
(a) Human Performance Complexity

Complexity Measure	r_s^*
metricalonormpk-MUS-44	0.687
keith	0.687
wnbd2	0.681
fitch	0.672
lhl	0.672
smith	0.672
metricalonorm	0.637
metricalponormpk-MUS-44	0.602
wnbd4	0.565
lz	0.562
offbeatness	0.532

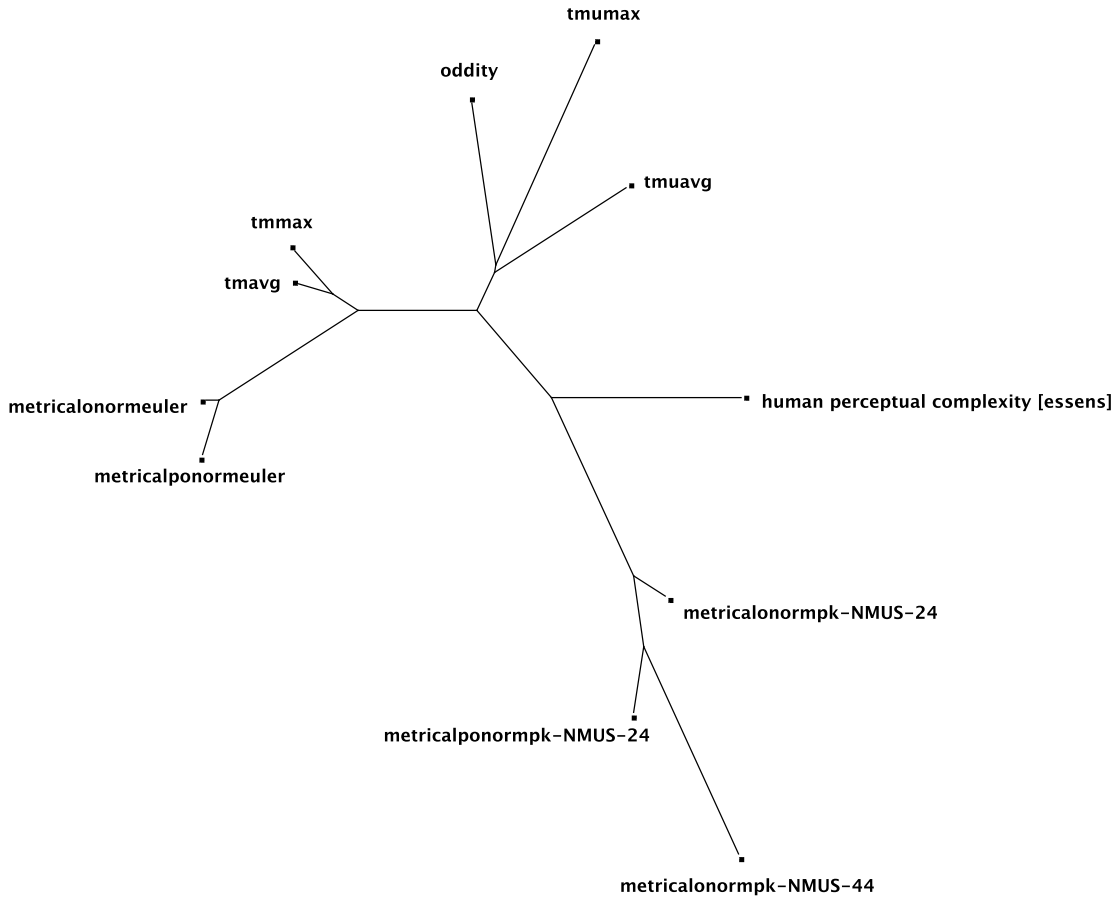
(b) Human Perceptual Complexity

Complexity Measure	r_s^*
oddity	0.397
metricalonormpk-NMUS-44	0.392
metricalonormeuler	0.382
metricalponormeuler	0.353
metricalonormpk-NMUS-24	0.344
tmumax	0.339
metricalponormpk-NMUS-24	0.315
tmavg	0.298
metricalponormpk-NMUS-44	0.273
tmuavg	0.205
tmmax	0.163

Regarding Essens' Human Perceptual Complexity, Table 6.2(b) reveals that the correlation of the measures is quite low, where the best correlation being $r_s^* = 0.397$ ($p \leq 0.05$), obtained by *oddity*. From the table, one can see that the other measures do not retain much



(a) BIONJ phylogenetic tree of the bottom of the horizontal cut, least-squares fit 97.7.



(b) BIONJ phylogenetic tree of the right of the vertical cut, least-squares fit 96.2.

Figure 6.4: *BIONJ phylogenetic trees of the complexity measures in the bottom of the horizontal cut along with Essens' Human Performance Complexity, and the tree of the right of the vertical cut along with Essens' Human Perceptual Complexity, see Figure 6.3 which shows the cuts.*

correlation to the human-based perceptual complexity.

However, even if the coefficients do not indicate strong correlation, the phylogenetic tree yields useful information. Consider Figure 6.4(b) which shows the result from the vertical cut of the tree in Figure 6.3. One noteworthy aspect of the sub-tree in Figure 6.4(b) is the cluster of the three Metrical Complexity measures using the Palmer and Krumhansl non-musicians $2/4$ and $4/4$ weighting schemes. Note that in the psychological experiment conducted by Essens, only 25% of the participants were musically trained [48]. Compare this to the 100% of participants musically trained in Shmulevich and Povel's study [146]. Essens' 25% seems rather small in comparison. Therefore, it is noteworthy that the weights which, performed best compared to the human-based perceptual complexity, were those of the non-musicians.

6.1.3 Fitch and Rosenfeld

The most current psychological study involving the complexity of rhythm is by Fitch and Rosenfeld [57] in 2007. Their work measures human-based performance complexity and metrical complexity. Fitch and Rosenfeld measured performance complexity in terms of a play-back error, and so we term their measure to be *Fitch and Rosenfeld's Human Performance Complexity*. As for the metrical complexity, two methods were used. Therefore we have *Fitch and Rosenfeld's Human Metrical Complexity (Beat-Tapping)* and *Fitch and Rosenfeld's Human Metrical Complexity (Number of Resets)*.

Consider Figure 6.5 depicting a phylogenetic tree the results. The BIONJ algorithm was used to construct the tree in the program SplitsTree. The cut in the tree separates the two measures of human-based metrical complexity from the human-based performance complexity. This was done because of the dense cluster near the human complexities, and differences are hard to see. Figure 6.6(a), shows the measures on the right of the cut.

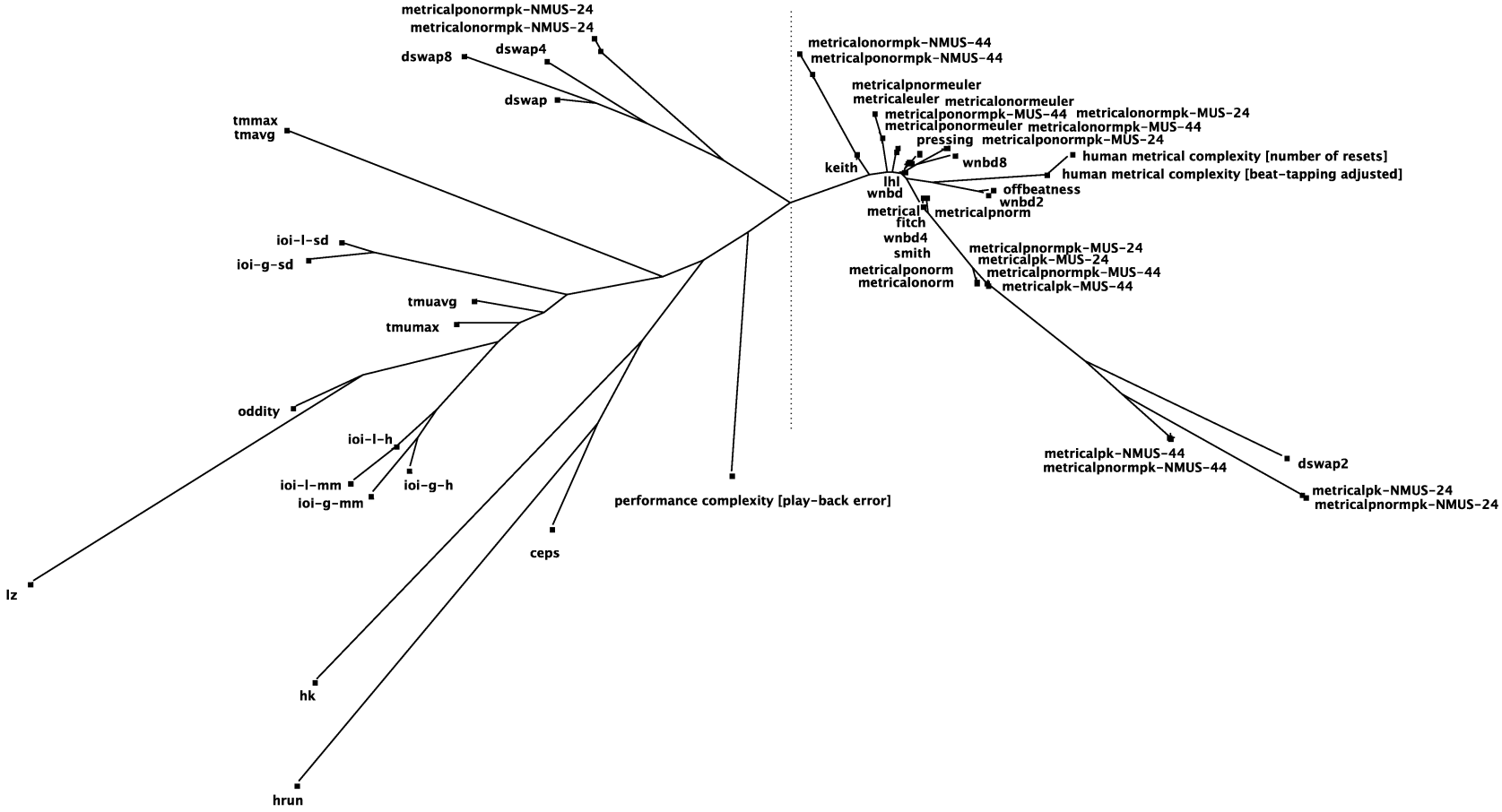


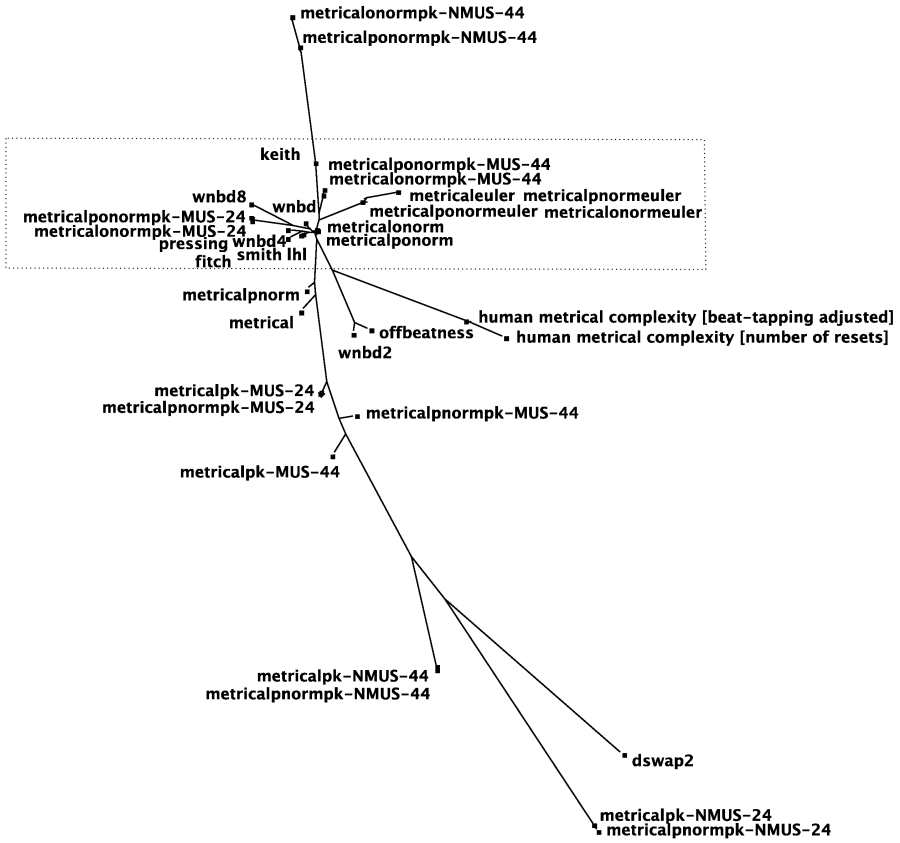
Figure 6.5: *BIONJ phylogenetic tree of the complexity measures, and Fitch and Rosenfeld's Human Metrical Complexity (Beat Tapping Adj), Human Metrical Complexity (Number of Resets), and Human Performance Complexity, least-squares fit 96.1.*

However, before looking at the sub-trees created from the cut, let us look at Figure 6.5. Since there are human measures of metrical and performance complexity, consider the computational measures which have a close relationship to Fitch and Rosenfeld's Human Metrical Complexity (Beat-Tapping) and Fitch and Rosenfeld's Human Metrical Complexity (Number of Resets). Surrounding the human measure, there is a large cluster. The cluster consists of those measures which use the metrical hierarchy. For example, the Metrical Complexity measure with normalization, and the versions which use Palmer and Krumhansl's weights, along with the Euler weights can be seen in this cluster. Moreover, the Longuet-Higgins and Lee implementations: *lhl*, *fitch*, and *smith* are also present in the cluster. We also see familiar top-performers such as *keith*, *pressing*, and *wnbd*. However, from this dense picture it is difficult to see some of the more subtle relationships in the cluster. Now turn to Figure 6.6(a), which depicts the measure to the right of the cut from Figure 6.5.

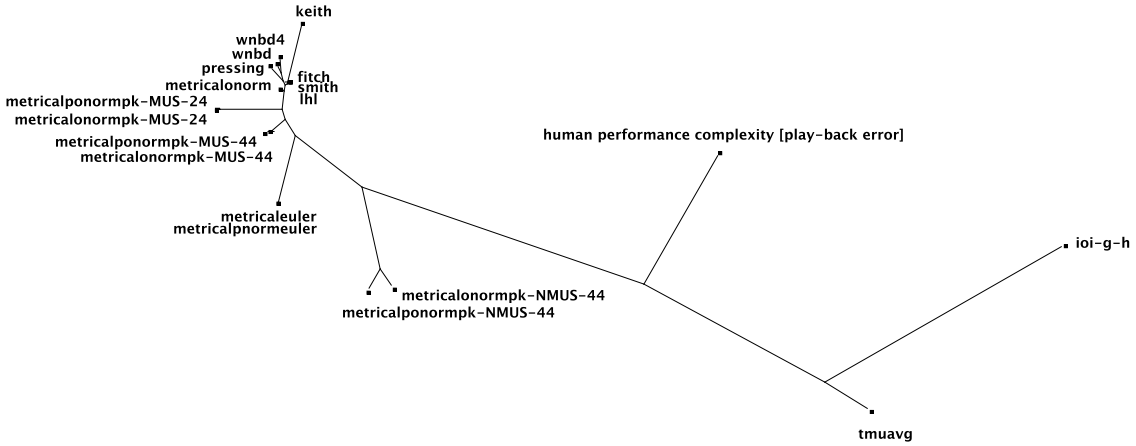
In Figure 6.6(a), note that dotted box. The complexity measures inside the box have very high intercorrelation: $r_s^* > 0.900$ with p-values at the $p \leq 0.001$ level for each. Therefore, we have a strong, significant relationship between these measures, and these measures also have high performance compared to Fitch and Rosenfeld's Human Metrical Complexity (Beat-Tapping and Number of Resets). For those complexity measures which are outside the box, a noteworthy separation is present. Towards the top and bottom of the tree, we see the Metrical Complexity measure using variants of the non-musicians weighting scheme, whereas the weighting scheme of the musicians is inside the tight cluster of the dotted box. For the precise correlation values, consult Table 6.3 (a) and (b) which show all measures retaining a coefficient $r_s^* > 0.600$ and significant p-values of $p < 0.001$. Note that we raised the r_s^* criterion to 0.600 from 0.500 because of the large number of measures which performed well with respect to the human-based metrical complexity measures.

Regarding Fitch and Rosenfeld's Human Performance Complexity, the picture is not so clear. Looking back to Figure 6.5, we can see that Human Performance Complexity is rather far from the computational measures of complexity. Therefore, we cannot conveniently cut the tree to see those computational measures with a strong relationship to the human measure. However, in order to gain some sense of the relationship, we picked measures using the criteria $r_s^* > 0.500$ and $p < 0.01$. This gave a handful of measures which are shown in Figure 6.6(b).

In Figure 6.6(b), we can more clearly see that Fitch and Rosenfeld's Human Performance Complexity resides far away from the measures which perform best. Moreover, the highest correlation is that of the *tmuavg* measure ($r_s^* = 0.613$), which is quite surprising since this measure has retained low accuracy regarding other human measures (c.f.



(a) BIONJ phylogenetic tree of the complexity measures in the right of the cut along with Fitch and Rosenfeld's Human Metrical Complexity (Beat-Tapping Adj) and (Number of Resets), least-squares fit 97.7.



(b) BIONJ phylogenetic tree of the complexity measures measures closest to Fitch and Rosenfeld's Human Performance Complexity, least-squares fit 99.4.

Figure 6.6: *BIONJ phylogenetic tree of the complexity measures in the right of the cut in (a) from Figure 6.5 along with the two human-based metrical complexities (Beat-Tapping and Reset Number), and a tree of those measures closest to the human-based performance complexity, since a clear cut was not present.*

Figures 6.1 and 6.3). In fact, the top three closest computational measures to the human-based performance complexity are each surprising in their own right. As mentioned, we have *tmuavg* as the closest, then *metricalonormpk-NMUS-44* which uses the non-musician's weighting scheme, and finally *ioi-g-h* which uses an interonset interval histogram and calculates the entropy of each normalized bin. Also, the best measures of the previous psychological data sets: *lhl, fitch, smith, wnbd*, and *keith*, have moderate correlation ($r_s^* \approx 0.500$ with $p \leq 0.01$), which is inconsistent with previous results.

6.1.4 Summary

The first set of results concern evaluating and validating the computational rhythmic complexity measures from psychological studies on three different data sets. The first data set is from Povel and Essens [119] in 1985, which studied measuring human-based performance complexity, this same data was then later used by Shmulevich and Povel [146] in 2000, to study human-based perceptual complexity. The second data set is from Essens [48] in 1995, where human-based performance and perceptual complexity were studied. Finally the third data set is from Fitch and Rosenfeld [57] in 2007, where human-based performance complexity and two types of human-based metrical complexity were studied. Below we present the findings by each type of human measure.

Human Performance Complexity

All three data sets gave a human measure for rhythm performance. The best measures with respect to their correlation values are listed in Table 6.4 for each data set.

In Table 6.4, there is little in common between all three data sets. The only complexity measure which remains in the top-five across the three is a version of the Metrical Complexity measure which uses the Palmer and Krumhansl weighting scheme. However, it should be noted that in Povel and Essens' data, as seen in Table 6.4(a), the weighting scheme of the time signature $2/4$ was preferred, and in Fitch and Rosenfeld's data (Table 6.4(c)) the non-musicians scheme was preferred.

Measures which did not correlate well with human-based performance complexity, but still had significant coefficients are the Metrical Complexity measures using Palmer and Krumhansl's metrical weighting of non-musicians for both Povel and Essens ($r_s^* \approx 0.400$), and Essens ($r_s^* = -0.400$). For Fitch and Rosenfeld, Tanguiane's Unconstrained Complexity measure ($r_s^* = 0.381$) was least correlated followed by the measure based on finding the tallest bin in the local IOI histogram, *ioi-l-mm* with $r_s^* = 0.387$.

Overall, there does not appear to be a clear best rhythm complexity measure for human-based performance complexity. We can say that measures such as the Longuet-Higgins and Lee yield high results in the Povel and Essens data set and Essens data set. The results of Fitch and Rosenfeld's data does not present a separation for the best measure.

Table 6.3: *In (a), the measures correlate to the beat tapping adjusted, with* $r_s^* > 0.600$ *and* $p \leq 0.001$. *In (b), the measures correlate to the number of resets, with* $r_s^* > 0.600$ *and* $p \leq 0.001$. *In (c), the measures correlate to human-based performance having* $r_s^* > 0.500$ *and* $p \leq 0.01$.

(a) Beat-Tapping Adj

Complexity Measure	r_s^*
metricalponorm	0.834
metricalonorm	0.826
metricalponormpk-MUS-44	0.813
metricalpnorm	0.812
metrical	0.812
wnbd	0.808
wnbd2	0.807
lhl	0.802
smith	0.802
metricalonormpk-MUS-44	0.801
offbeatness	0.800
wnbd4	0.797
fitch	0.793
syncopation	0.793
metricalponormeuler	0.792
metricalponormpk-MUS-24	0.789
metricalonormeuler	0.781
metricalonormpk-MUS-24	0.779
keith	0.766
pressing	0.764
wnbd8	0.761
metricalpnormeuler	0.745
metricaleuler	0.745
metricalpnormpk-MUS-24	0.725
metricalpk-MUS-24	0.724
metricalpk-MUS-44	0.722
metricalpnormpk-MUS-44	0.722
metricalponormpk-NMUS-44	0.718
metricalonormpk-NMUS-44	0.670

(b) Number of Resets

Complexity Measure	r_s^*
metricalponorm	0.793
metricalonorm	0.786
metrical	0.781
metricalpnorm	0.781
wnbd	0.777
wnbd4	0.765
wnbd2	0.762
metricalponormpk-MUS-44	0.760
lhl	0.757
smith	0.757
syncopation	0.748
fitch	0.748
metricalonormpk-MUS-44	0.747
metricalponormeuler	0.744
offbeatness	0.743
metricalponormpk-MUS-24	0.734
keith	0.731
metricalonormeuler	0.730
wnbd8	0.730
metricalonormpk-MUS-24	0.724
pressing	0.723
metricalpnormpk-MUS-24	0.697
metricalpk-MUS-24	0.696
metricalpk-MUS-44	0.694
metricalpnormpk-MUS-44	0.693
metricalpnormeuler	0.688
metricaleuler	0.688
metricalponormpk-NMUS-44	0.681
metricalonormpk-NMUS-44	0.620

(c) Performance

Complexity Measure	r_s^*	**Complexity Measure**	r_s^*
tmuavg	0.613	fitch	0.521
metricalonormpk-NMUS-44	0.596	lhl	0.520
ioi-g-h	0.573	metricalponorm	0.520
metricalponormpk-MUS-44	0.551	smith	0.520
metricaleuler	0.548	metricalponormpk-MUS-24	0.519
metricalpnormeuler	0.548	metricalonormpk-MUS-24	0.516
metricalonormpk-MUS-44	0.543	metricalonorm	0.515
wnbd	0.534	keith	0.511
metricalponormpk-NMUS-44	0.527	pressing	0.509
wnbd4	0.526		

Table 6.4: *In (a), we have the top-five individual complexity measures with respect to human-based performance complexity from Povel and Essens' data set. In (b), we have the top-five individual complexity measures with respect to human-based performance complexity from Essens' data set. In (c), we have the top-five individual complexity measures with respect to human-based performance complexity from Fitch and Rosenfeld's data set.*

(a) Povel and Essens

Complexity Measure	r_s^*
smith	0.800
fitch/lhl	0.787
wnbd/wnbd4	0.747/0.700
metricalpk-MUS-24	0.687
metrical	0.682

(b) Essens

Complexity Measure	r_s^*
metricalonormpk-MUS-44/keith	0.687
wnbd2	0.681
lhl/fitch/smith	0.672
metricalonorm	0.637
metricalponormpk-MUS-44	0.602

(c) Fitch and Rosenfeld

Complexity Measure	r_s^*
tmuavg	0.613
metricalonormpk-NMUS-44	0.596
ioi-g-h	0.573
metricalponormpk-MUS-44	0.551
metricaleuler	0.548

Human Perceptual Complexity

For the human-based perceptual complexity, there are two data sets which yield human measurements: Shmulevich and Povel in 2000 and Essens in 1995. Table 6.5 below includes the complexity measures which have the highest correlation. However, as can be seen in Table 6.5(e), even the best correlations are not very strong.

(d) Shmulevich and Povel

Complexity Measure	r_s^*
lhl/fitch	0.755
wnbd/wnbd4	0.738/0.690
smith	0.737
metrical	0.694
metricalponormpk-MUS-24	0.682

(e) Essens

Complexity Measure	r_s^*
oddity	0.397
metricalonormpk-NMUS-44/24	0.392/0.344
metricalonormeuler	0.382
metricalponormeuler	0.353
tmumax	0.339

Table 6.5: *In (d), we have the top-five individual complexity measures with respect to human-based perceptual complexity from Povel and Essens' data set and Shmulevich and Povel's experimentation. In (e), we have the top-five individual complexity measures with respect to human-based performance complexity from Essens' data set.*

From Table 6.5(d) alone, those measures based on the metrical hierarchy such as the Longuet-Higgins and Lee measure and the Metrical Complexity measure have a high correlation to Shmulevich and Povel's Human Perceptual Complexity with $r_s^* = 0.755$. The WNBD measure also gives a high correlation. While no complexity measure gives a high

correlation to Essens' Human Perceptual Complexity, as seen in Table 6.5(d), of the measures with the highest correlation, variants of the Metrical Complexity measure using alternate weighting schemes are present. This result agrees with Shmulevich and Povel's Human Perceptual Complexity in that the measures based on the metrical hierarchy resulted as top performers.

Regarding the complexity measures which did not perform well, we shall only mention those for Shmulevich and Povel's Human Perceptual Complexity. This is because the Essens' data did not yield high correlations. However, with respect to Shmulevich and Povel, Tanguiane's (Maximum) measure ($r_s^* = 0.331$), Rhythmic Oddity ($r_s^* = 0.342$), and the Directed Swap (Four Beat) Distance ($r_s^* = 0.346$), all had low correlations. Note that their p-values were $p < 0.05$.

Human Metrical Complexity

Fitch and Rosenfeld's study [57] provides a human measure which has not been seen in the literature thus far: a measure for human metrical complexity. Table 6.6 shows the highest correlations regarding Fitch and Rosenfeld's Human Metrical (Beat-Tapping and Number of Resets) Complexity measures.

Table 6.6: *In (a) and (b), we have the top-five individual complexity measures with respect to each type of human-based metrical complexity from Fitch and Rosenfeld's data set.*

(a) Fitch and Rosenfeld, Beat-Tapping

Complexity Measure	r_s^*
metricalponorm	0.834
metricalonorm	0.826
metricalponormpk-MUS-44	0.813
metricalpnorm/metrical	0.812
wnbd	0.808

(b) Fitch and Rosenfeld, Number of Resets

Complexity Measure	r_s^*
metricalponorm	0.793
metricalonorm	0.786
metrical	0.781
metricalpnorm	0.781
wnbd	0.777

In both Tables 6.6(a) and 6.6(b), the most highly correlated complexity measures are variants on the Metrical Complexity measure. As can be seen from the tables, not only do the Metrical Complexity measure variants dominate the top-five results, but the correlations are the highest we've seen so far with $r_s^* = 0.834$. In addition to this, we have the WNBD's results just under the Metrical Complexity's results. One notable difference between the two human-based metrical complexity comparisons is that in the case of Beat-Tapping, the weighting scheme of Palmer and Krumhansl's musicians in $4/4$ time is present in the list, whereas this is not so in the second case of the Number of Resets.

6.2 Cultural

Here we discuss musical rhythm complexity in terms of two cultures: African and Indian. Below we first detail the results of using African *Timelines*, and second we detail

results of using Indian *Decitalas* and North Indian *Talas*. Phylogenetic trees are use to visualize the results. From the results, we aim to show which measures of complexity are robust between rhythms with a different number of pulses (12 or 16), and also between different cultures.

In order to determine the robustness of the complexity measures among and between cultures, we investigate how the phylogenetic trees change. For example, we can look at the correlations between all pairs of measures when using African *Timelines* of 12 pulses and then look at correlations between all pairs of measures using African *Timelines* of 16 pulses. However, since there are such a large number of pairs, one way to see a change between the trees is to look at how the clustering between complexity measures changes. The hypothesis is that if a complexity measure is robust, then the intercorrelation will be stable. That is, robust complexity measures will cluster similarly in one phylogenetic tree compared to another, showing consistent (i.e., stable) intercorrelation.

We test this hypothesis by looking for complexity measures whose Spearman rank correlation coefficients are not stable between phylogenetic trees. Therefore, we can readily pick-out the unstable measures, and assume the others attain relatively stable performance. A method for determining which measures are not stable is to plot the change in coefficient values with a *boxplot* [175], as introduced by J. W. Tukey in 1977. More specifically, the data we plot with this method are the Spearman rank correlation coefficients between complexity measures according to one data set, subtracted from the coefficients of the same pairs, in a second data set. Additionally, since we are only interested in the magnitude of change in the correlation coefficient between pairs of measures, we take the absolute value of the calculated difference.

For example, consider the correlation between the Longuet-Higgins and Lee Complexity measure, and Pressing's Cognitive Complexity measure on African *Timelines* with 12 pulses. Let this correlation be α_{rs}. Now, consider the correlation between the same two complexity measures, but this time, on African *Timelines* with 16 pulses. Let this correlation be β_{rs}. We define the absolute value of the change in correlation by Equation (6.1) as follows. The value $|\Delta_{rs}|$ is found for all pairs of measures between two sets of correlation values. It is these values, Δ_{rs}, which we use for the boxplot.

$$|\Delta_{rs}| = |\alpha_{rs} - \beta_{rs}| \tag{6.1}$$

The goal of the boxplot visualization is to determine whether there is a significant change in correlation of pairs of measures between the two trees. We consider a significant change to represent an outlier in the data. Thus, by using the boxplot, we may discover outliers without making any assumptions about the underlying distribution of the data [175].

Below, we discuss outliers found by the boxplots; i.e., those complexity measures which are not robust. We do this for rhythms within the same culture, and also between different cultures. Note that all boxplots are made using MATLAB® [163].

6.2.1 Intra-Cultural Comparison

Here we make a comparison between African *Timelines* with 12 pulses to those with 16 pulses, and then we show the Indian *Decitalas* with 12 pulses compared to those with 16 pulses.

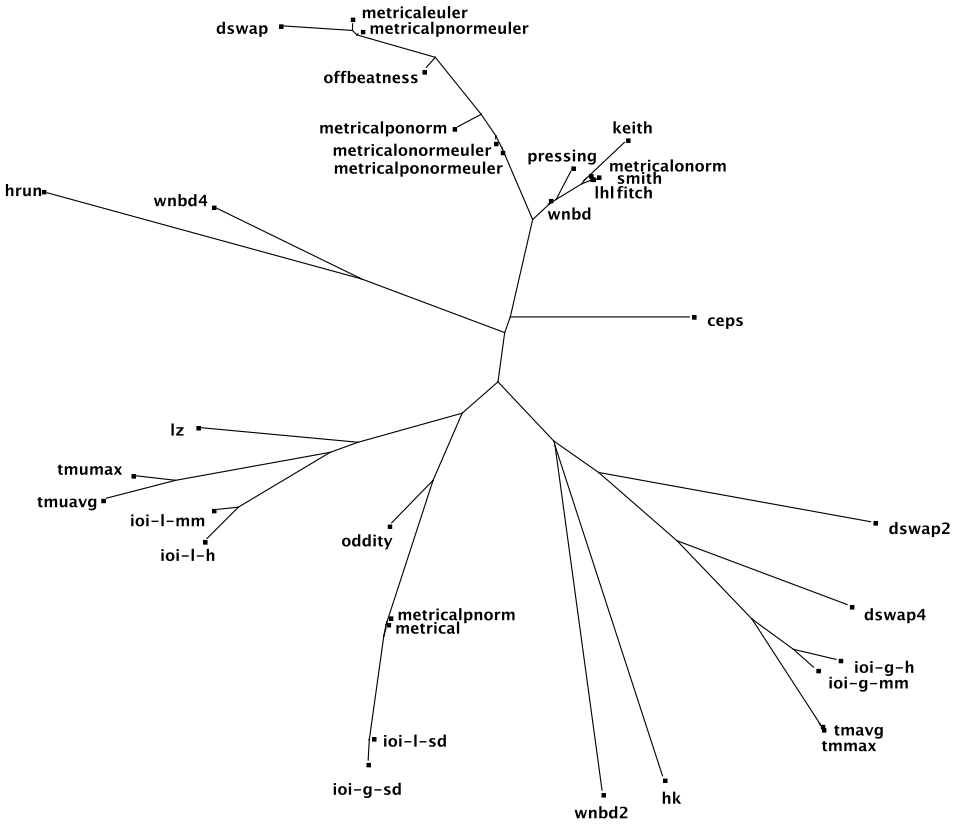
African *Timelines*

The results here show the robustness (or stability) of the complexity measures between African *Timelines* with 12 pulses and African *Timelines* with 16 pulses. Thus, the robustness criterion we are testing is the length (in terms of pulses) of the rhythms. Below, we use the boxplot to visualize any outlier complexity measures which we consider to not be robust with respect to the length of the rhythm. However, first consider two phylogenetic trees shown in Figure 6.7.

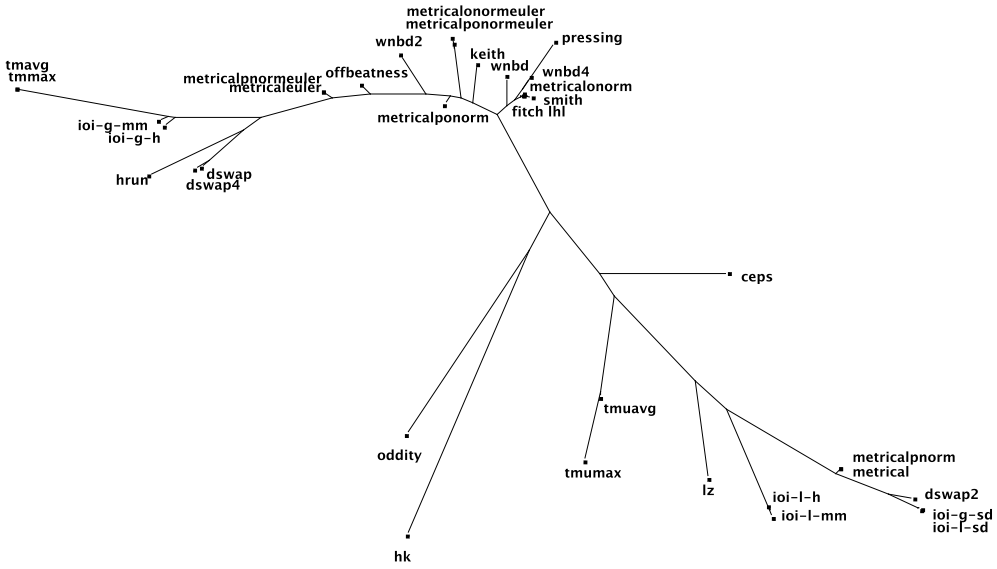
In Figure 6.7(a), we have a phylogenetic tree depicting the relationships among complexity measures on African *Timelines* of 12 pulses. In Figure 6.7(b), we have a tree showing those same measures on African *Timelines* of 16 pulses. Note that we chose measures which are applicable to rhythms which have 12 or 16 pulses. Between the figures, one may notice that there are clusters which are similar and different. For example, one cluster common to both the trees is the one containing the measures: *fitch, lhl, metricalonorm*, and *smith*. One cluster which is different is in Figure 6.7(a) we have *dswap2* is moderately correlated with *dswap4*; however, in Figure 6.7(b) this is clearly not the case as *dswap2* and *dswap4* are almost on opposites ends of the tree. Differences between the trees, such as the one just discussed, are what we aim to discover, because we hypothesize that those differences reveal measures that are not robust. In order to visualize such outliers, consider the following boxplot in Figure 6.8.

The boxplot shows a box which represents the interquartile range of the distribution. The line which vertically cuts the box represents the midpoint of the distribution. The horizontal dotted lines which end with a vertical line are the whiskers of the diagram and mark the cutoffs where any point beyond the vertical lines are considered outliers. It is those points which are further right (or left) than the vertical lines extending from the dotted horizontal lines which we are interested in. Figure 6.8 shows that any point; i.e., any $|\Delta_{rs}|$ which is greater than 0.680 is considered to be an outlier. Now, we use this as a cutoff and find the pairs of complexity measures which have a change in coefficient greater than 0.680. For convenience, these pairs are listed in Table 6.7.

The complexity measures listed in Table 6.7 are considered to be the outliers for com-



(a) BIONJ phylogenetic tree of the complexity measures on African *Timelines* with 12 pulses, least-squares fit 93.2.



(b) BIONJ phylogenetic tree of the complexity measures on African *Timelines* with 16 pulses, least-squares fit 92.4.

Figure 6.7: *BIONJ phylogenetic trees of the complexity measures which may be used on African* Timelines *with 12 and 16 pulses.*

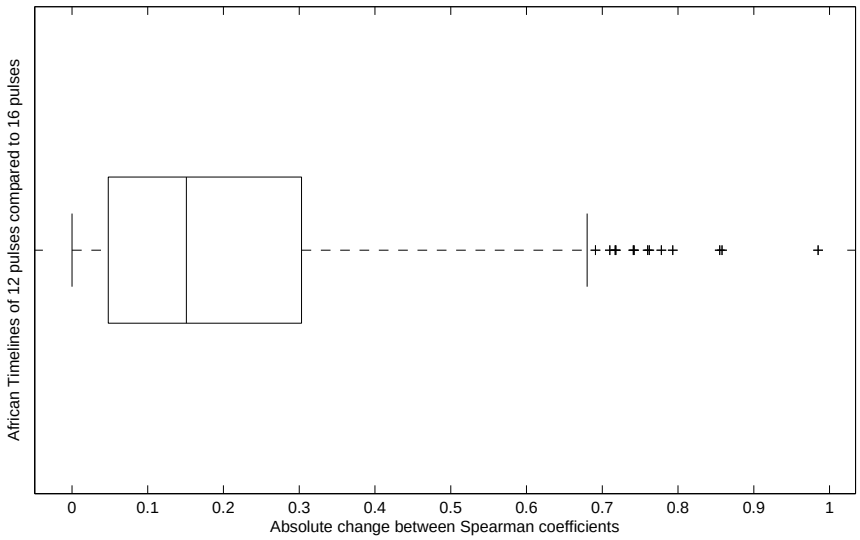


Figure 6.8: *Boxplot of the absolute value of the change ($|\Delta_{rs}|$) between corresponding pairs of Spearman rank correlation coefficients of complexity measures from the African* Timelines *with 12 pulses to the African* Timelines *with 16 pulses, outlier cutoff* $|\Delta_{rs}| > 0.680$.

Table 6.7: *Outliers (as indicated by the boxplot in Figure 6.8) with an absolute-value change,* $|\Delta_{rs}| > 0.680$*, in Spearman rank correlation coefficient going from the African* Timelines *with 12 pulses (Figure 6.7(a)) and African* Timelines *with 16 pulses (Figure 6.7(b)).*

Complexity Measure Pair		$\lvert\Delta_{rs}\rvert$
dswap4	dswap2	0.985
wnbd2	offbeatness	0.858
dswap4	hrun	0.855
wnbd4	keith	0.793
wnbd2	keith	0.778
wnbd2	metricalponorm	0.762
wnbd2	smith	0.760
wnbd2	metricalonorm	0.742
ioi-l-sd	dswap2	0.741
wnbd2	lhl	0.717
wnbd2	fitch	0.710
wnbd2	metricalonormeuler	0.691

paring the intercorrelation of the measures on African *Timelines* with 12 pulses to the intercorrelation for African *Timelines* with 16 pulses. Thus, we see that the difference between the *dswap2* and *dswap4* is indicated to be the largest outlier, which we noted to be a distinct change from investigating the phylogenetic trees. Another noteworthy difference between the trees is the *wnbd2* performance. We can see from the table that this measure has significant differences compared to eight different measures. Also, note that *keith* is listed twice in the table. Thus, we have *dswap2, dswap4, keith*, and *wnbd2* listed in more than one outlier pair.

Indian *Decitalas*

As in the above, we test for robustness of the complexity measures for Indian *Decitalas* with 12 pulses compared to those with 16 pulses. Again our criterion is the length of the rhythmic pattern; however, we use Indian *Decitalas*. Consider Figure 6.9 which shows the relationships of the measures using *Decitalas* with 12 pulses, seen in Figure 6.9(a), and *Decitalas* with 16 pulses, shown in Figure 6.9(a).

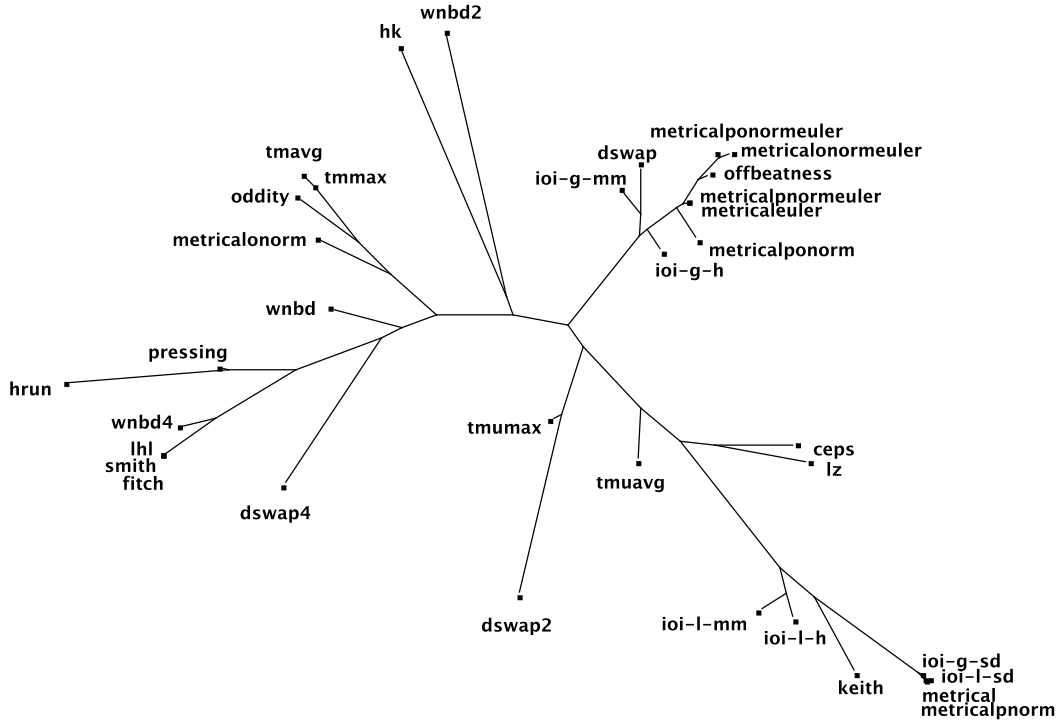
Visual inspection for measures, which may not be robust, shows a few possibilities; however, we turn to the boxplot in order to discover the true outliers among the differences between coefficients. Consider Figure 6.10 which shows a boxplot of the the distribution of the $|\Delta_{rs}|$ values.

In the boxplot of Figure 6.10, we see that the cutoff for outliers is $|\Delta_{rs}| > 1.217$, and we see a few values which are above that cutoff. These are marked by a $+$ symbol. If we lookup these values, we find the following pairs of complexity measures along with their $|\Delta_{rs}|$ values, as shown in Table 6.8 below.

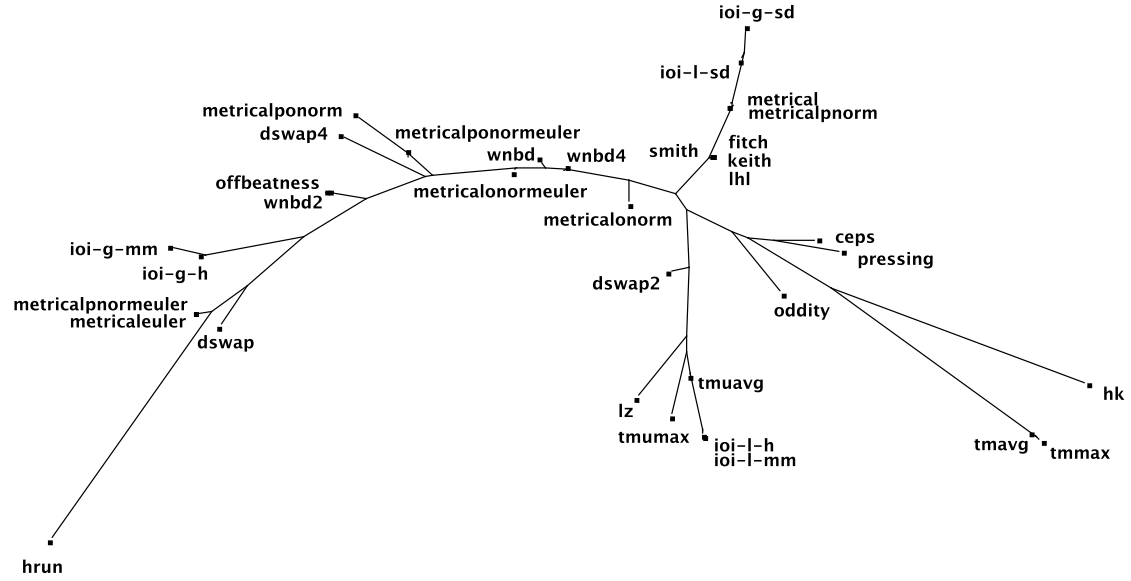
Table 6.8: *Outliers (as indicated by the boxplot in Figure 6.10) with an absolute-value change,* $|\Delta_{rs}| > 1.217$*, in Spearman rank correlation coefficient going from the Indian* Decitalas *with 12 pulses (Figure 6.9(a)) and Indian* Decitalas *with 16 pulses (Figure 6.9(b)).*

Complexity Measure Pair		$\lvert\Delta_{rs}\rvert$
keith	wnbd4	1.539
keith	smith	1.281
keith	lhl	1.280
keith	fitch	1.280
tmuavg	wnbd4	1.256
keith	wnbd	1.246

From Table 6.8, we can see that *keith* and *wnbd4* are the most frequent measures among the outlier pairs. Looking back at the trees, we can clearly see in Figure 6.9(b) that *keith* is very close to the cluster: *fitch, lhl*, and *smith*. However, looking at Figure 6.9(a), *keith* is quite distant from that cluster.



(a) BIONJ phylogenetic tree of the complexity measures on Indian *Decitalas* with 12 pulses, least-squares fit 90.1.



(b) BIONJ phylogenetic tree of the complexity measures on Indian *Decitalas* with 16 pulses, least-squares fit 89.6.

Figure 6.9: *BIONJ phylogenetic trees of the complexity measures which may be used on Indian* Decitalas *with 12 and 16 pulses.*

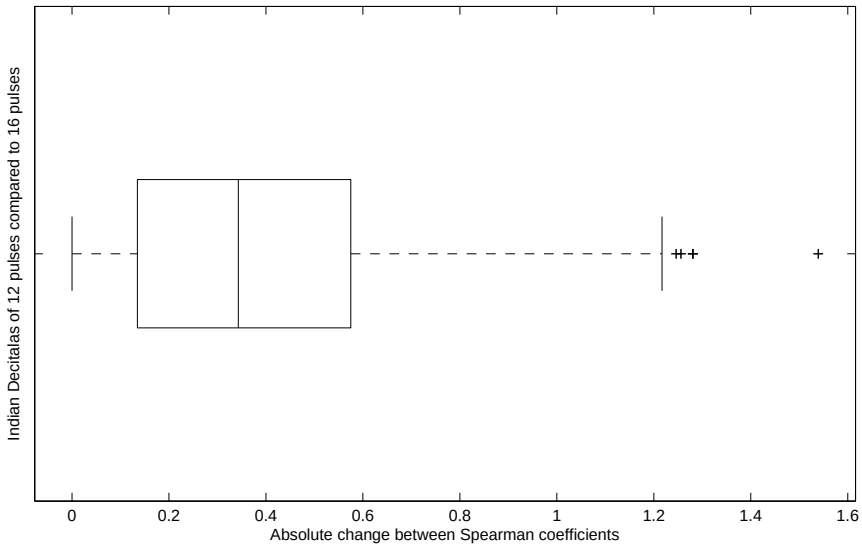


Figure 6.10: *Boxplot of the absolute value of the change ($|\Delta_{rs}|$) between corresponding pairs of Spearman rank correlation coefficients of complexity measures from the Indian* Decitalas *with 12 pulses to the Indian* Decitalas *with 16 pulses, outlier cutoff* $|\Delta_{rs}| > 1.217$.

6.2.2 Inter-Cultural Comparison

Now we validate the complexity measures based on their robustness when comparing data sets which have rhythms from different cultures. First, we compared African *Timelines* with 12 or 16 pulses to Indian *Decitalas* with 12 or 16 pulses. Second, we compare African *Timelines* with 12 or 16 pulses to all of the Indian *Decitalas*. Third, we compare the African *Timelines* with 12 or 16 pulses to all of the North Indian *Talas*.

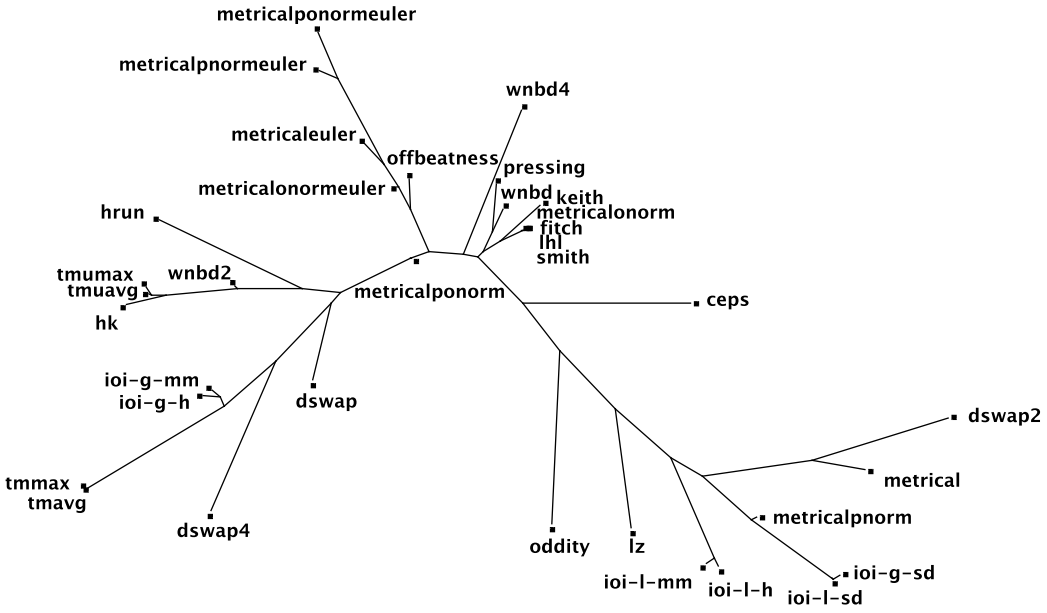
African (12,16)-*Timelines* vs. Indian (12,16)-*Decitalas*

The first comparison is between the African *Timelines* and Indian *Decitalas*, both sets having rhythms with 12 or 16 pulses. Figure 6.11 shows phylogenetic trees for the complexity measures on each set, the *Timelines* in Figure 6.11(a) and the *Decitalas* in Figure 6.11(b).

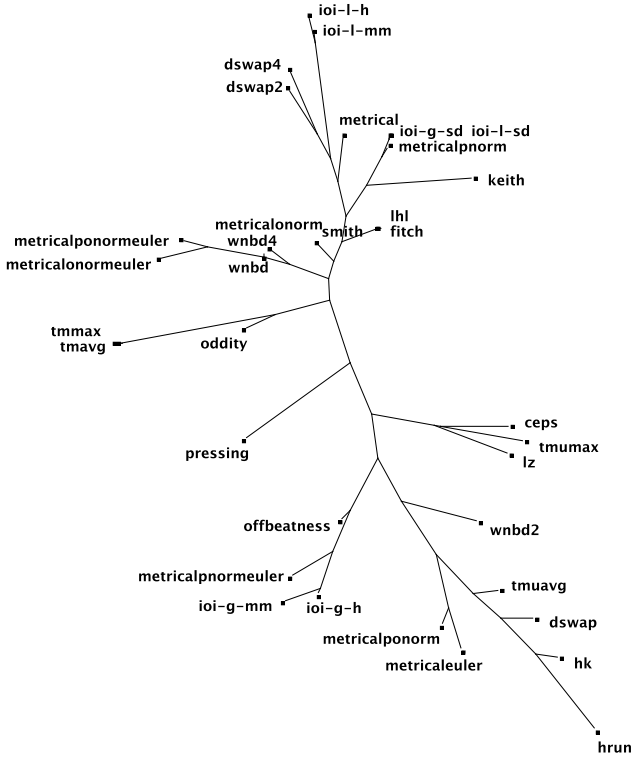
We are looking for measures where the coefficient between two complexity measures significantly changes from Figure 6.11(a) to Figure 6.11(b). We use a boxplot of the coefficients from pairs of measures in both trees in order to discover the cutoff for any outliers. Figure 6.12 below shows the boxplot.

Figure 6.12 shows that the cutoff $|\Delta_{rs}|$ is 0.874. We list all pairs above the cutoff, along with their $|\Delta_{rs}|$ values, in Table 6.9 below.

From Table 6.9, we can see that the complexity measures most frequent in the outlier pairs are: *hrun, ioi-g-mm, metricalponorm*, and *oddity*. As an example, consider the measure *metricalponorm*. In Figure 6.11(a) we can see that the measure resides near the middle



(a) BIONJ phylogenetic tree of the complexity measures on African *Timelines* with 12 or 16 pulses, least-squares fit 94.2.



(b) BIONJ phylogenetic tree of the complexity measures on Indian *Decitalas* with 12 or 16 pulses, least-squares fit 92.7.

Figure 6.11: *BIONJ phylogenetic trees of the complexity measures which may be used on African* Timelines *with 12 and 16 pulses, and Indian* Decitalas *with 12 or 16 pulses.*

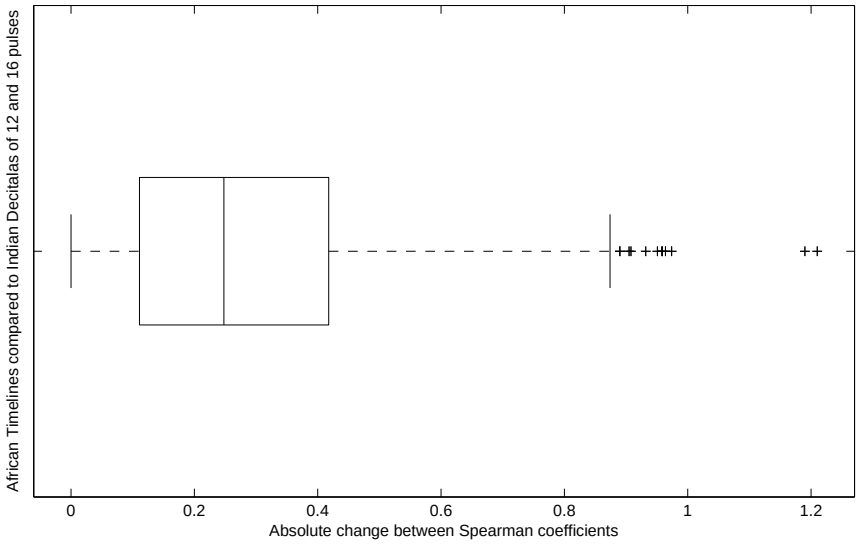


Figure 6.12: *Boxplot of the absolute value of the change ($|\Delta_{rs}|$) between corresponding pairs of Spearman rank correlation coefficients of complexity measures from the African* Timelines *with 12 or 16 pulses to the Indian* Decitalas *with 12 and 16 pulses, outlier cutoff* $|\Delta_{rs}| > 0.874$.

Table 6.9: *Outliers (as indicated by the boxplot in Figure 6.12) with an absolute-value change,* $|\Delta_{rs}| > 0.874$*, in Spearman rank correlation coefficient going from the African* Timelines *with 12 and 16 pulses (Figure 6.11(a)) and Indian* Decitalas *with 12 and 16 pulses (Figure 6.11(b)).*

Complexity Measure Pair		$\lvert\Delta_{rs}\rvert$
oddity	tmavg	1.210
oddity	tmmax	1.190
hrun	offbeatness	0.974
ioi-g-mm	tmavg	0.964
metricalponorm	lhl	0.959
metricalponorm	fitch	0.958
metricalponorm	smith	0.951
hrun	metricalonorm	0.932
ioi-g-mm	tmmax	0.908
hrun	wnbd4	0.905
dswap4	metrical	0.890
dswap2	wnbd4	0.890

of the tree close to measures like *lhl, fitch, smith*, etc. However, Figure 6.11(b) shows *metricalponorm* far from the *lhl* cluster, residing at the bottom of the tree.

African (12,16)-*Timelines* and Indian *Decitalas*

Here we compare the African *Timelines* to the Indian *Decitalas*. Note that this collection of *Decitalas* has an average number of pulses of 13.320 with a standard deviation of 6.748. Figure 6.13 shows the relationship among the complexity measures for the African *Timelines* of 12 and 16 pulses, Figure 6.13(a), and the Indian *Decitalas*, shown in Figure 6.13(b).

The two phylogenetic trees in Figure 6.13 reveal quite a few similarities among the clustering of the complexity measures. However, since we are looking for measures which may not be robust between the data sets, let us construct a boxplot of the $|\Delta_{rs}|$ values and look for any outliers. We can see the boxplot in Figure 6.14 below.

The boxplot in Figure 6.14 shows that there are very few outliers to the right of the cutoff of $|\Delta_{rs}| > 0.737$. By looking at the trees, this seems like a reasonable result, because many of the complexity measure do in fact remain in similar clusters. However, to find out which pairs are the outliers, consider Table 6.10.

Table 6.10: *Outliers (as indicated by the boxplot in Figure 6.14) with an absolute-value change, $|\Delta_{rs}| > 0.737$, in Spearman rank correlation coefficient going from the African* Timelines *with 12 and 16 pulses (Figure 6.13(a)) and Indian* Decitalas *with 12 and 16 pulses (Figure 6.13(b)).*

Complexity Measure Pair		$\lvert\Delta_{rs}\rvert$
offbeatness	metricalponorm	0.937
offbeatness	hrun	0.855
offbeatness	smith	0.829
offbeatness	lhl	0.826
offbeatness	fitch	0.825
offbeatness	keith	0.795

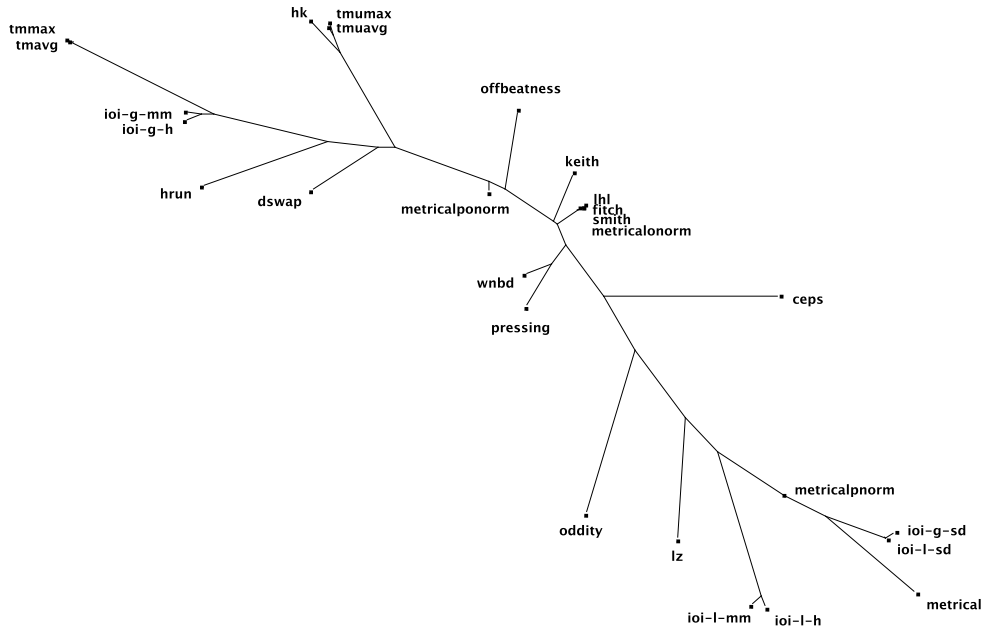
The outlier pairs listed in Table 6.10, pick out one measure in particular whose coefficients between the two data sets was significant. As can be seen, this measure is *offbeatness*. From Figure 6.13, the dramatic change in position of *offbeatness* can readily be seen.

African (12,16)-*Timelines* and North Indian *Talas*

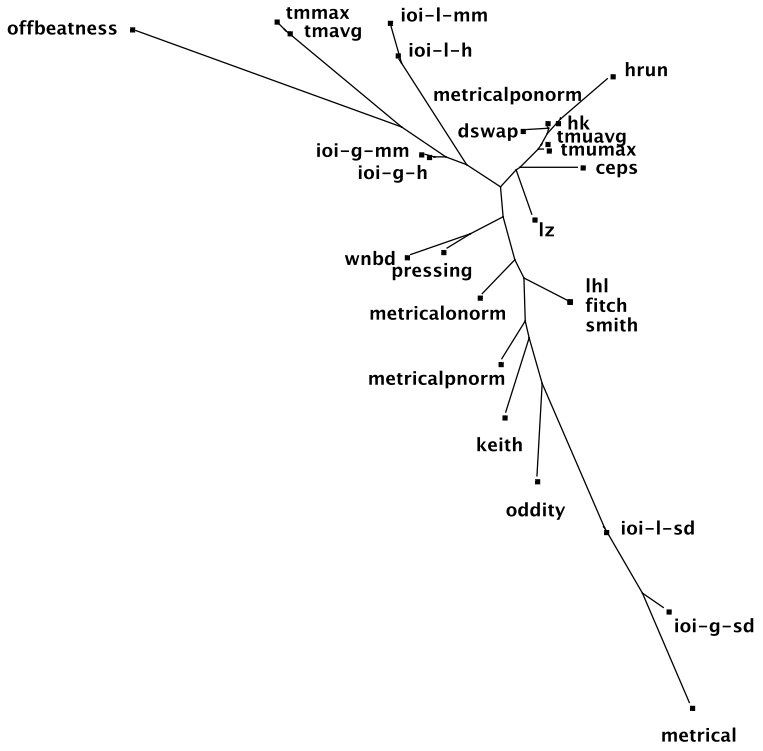
Another set of data are the North Indian *Talas*. Here, we compare the African *Timelines* with the North Indian *Talas*. Figure 6.15 shows the corresponding phylogenetic trees.

The phylogenetic trees in Figure 6.15 depict changes which are a bit more subtle. Perhaps the boxplot in Figure 6.16 will reveal outliers in the data. In fact, there are outliers, as show in the figure. Consider the boxplot in Figure 6.16 below.

Figure 6.16 shows outliers in the change in coefficients between the African *Timelines* and North Indian *Talas*. Table 6.11 lists theses outlier pairs.



(a) BIONJ phylogenetic tree of the complexity measures on African *Timelines* with 12 and 16 pulses, least-squares fit 94.0.



(b) BIONJ phylogenetic tree of the complexity measures on Indian *Decitalas*, least-squares fit 95.5.

Figure 6.13: *BIONJ phylogenetic trees of the complexity measures which may be used on African* Timelines *of 12 and 16 pulses and Indian* Decitalas*.*

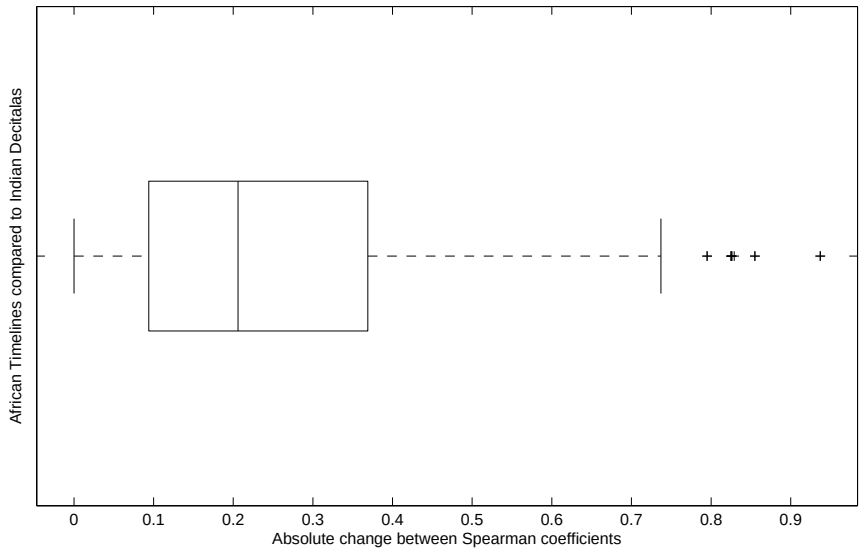
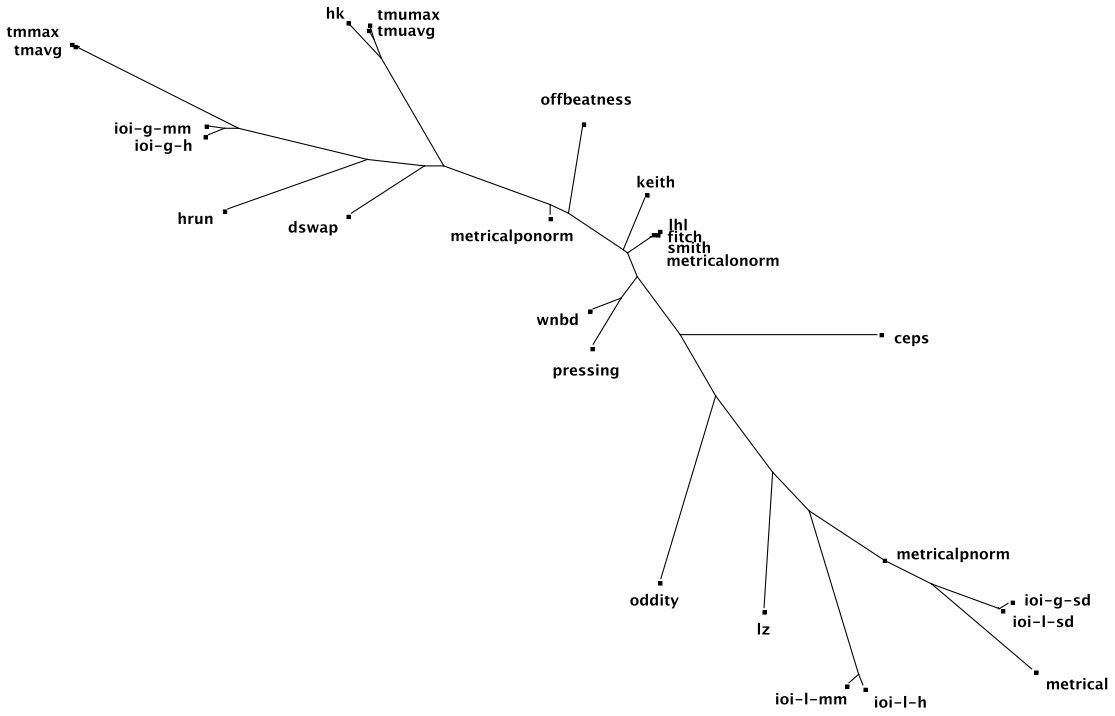


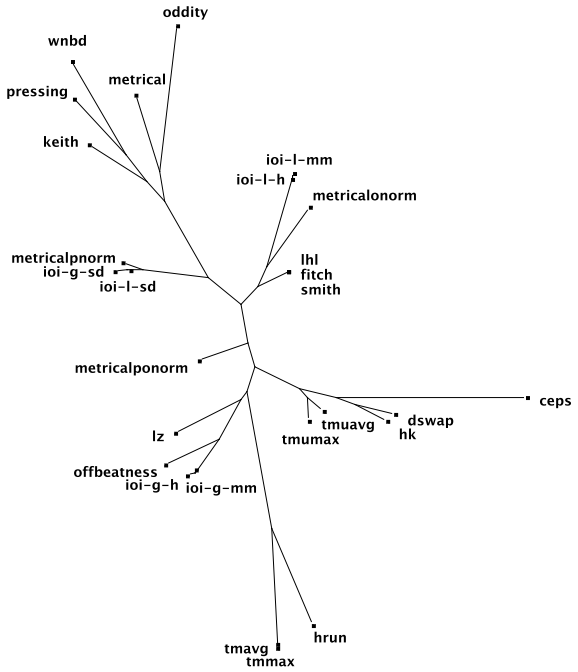
Figure 6.14: *Boxplot of the absolute value of the change ($|\Delta_{rs}|$) between corresponding pairs of Spearman rank correlation coefficients of complexity measures from the African* Timelines *with 12 and 16 pulses to the Indian* Decitalas*, outlier cutoff* $|\Delta_{rs}| > 0.737$.

Table 6.11: *Outliers (as indicated by the boxplot in Figure 6.16) with an absolute-value change,* $|\Delta_{rs}| > 0.825$*, in Spearman rank correlation coefficient going from the African* Timelines *with 12 and 16 pulses (Figure 6.15(a)) to all of the North Indian* Talas *(Figure 6.15(b)).*

Complexity Measure Pair		$\lvert\Delta_{rs}\rvert$
keith	metricalonorm	1.278
pressing	metricalonorm	1.205
wnbd	metricalonorm	1.020
wnbd	lhl	0.932
wnbd	fitch	0.931
wnbd	smith	0.922



(a) BIONJ phylogenetic tree of the complexity measures on African *Timelines* with 12 and 16 pulses, least-squares fit 94.0.



(b) BIONJ phylogenetic tree of the complexity measures on North Indian *Talas*, least-squares fit 93.2.

Figure 6.15: *BIONJ phylogenetic trees of the complexity measures which may be used on African* Timelines *with 12 and 16 pulses and North Indian* Talas*.*

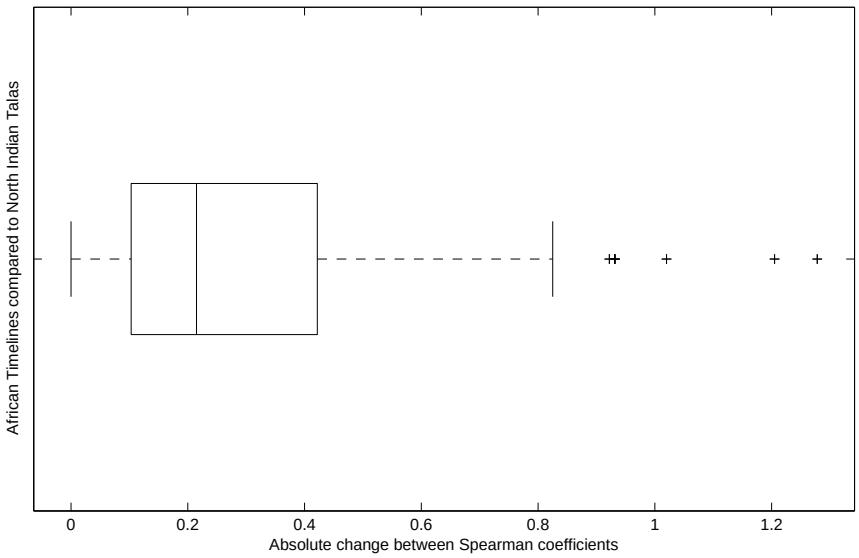


Figure 6.16: *Boxplot of the absolute value of the change ($|\Delta_{rs}|$) between corresponding pairs of Spearman rank correlation coefficients of complexity measures from the African* Timelines *with 12 and 16 pulses to the North Indian* Talas*, outlier cutoff* $|\Delta_{rs}| > 0.825$.

Table 6.11 shows two measures which show up frequently in the outlier pairs: *metricalonorm* and *wnbd*. From the trees in Figure 6.15, we can see that these measures are indeed outliers. In Figure 6.15(a), the *metricalonorm* measure is tucked into the cluster with *fitch, lhl*, and *smith*. However, *metricalonorm* gets pulled out of this cluster in Figure 6.15(b). Moreover, we have *wnbd* close to measures *fitch, lhl*, and *smith* in Figure 6.15(a), but in Figure 6.15(b), *wnbd* is far from those measures.

6.2.3 Summary

In Part 2 of the results, we observed the relationships between the complexity measures in two ways. First we looked at the relationship of the measures on data from the same culture where the number of pulses changed. This was shown for the African *Timelines* for rhythms with 12 and 16 pulses. Moreover, this was shown for the Indian *Decitalas* for rhythms with 12 and 16 pulses. Second we looked at the relationship of the complexity measures on rhythms between different cultures. Hence, we made three comparisons with the African *Timelines*: (1) the Indian *Decitalas* of 12 and 16 pulses, (2) the Indian *Decitalas*, and (3) the North Indian *Talas*. Below we provide a table which summarizes which complexity measures are in a pair of at least two of the outliers in the boxplots, and outlier tables for each respective comparison.

Intra-Cultural

Table 6.12 below consists of those complexity measures which are considered to be in an outlier pair for the robustness test within the African *Timelines* data set and the Indian *Decitalas* data set.

Table 6.12: *Summary table of complexity measures which are listed as outliers in more than one pair from the boxplots of the comparison of African* Timelines *with 12 pulses to African* Timelines *with 16 pulses (a), and Indian* Decitalas *with 12 pulses to Indian* Decitalas *with 16 pulses (b).*

(a) African *Timelines*

Complexity Measure
dswap2
dswap4
keith
wnbd2

(b) Indian *Decitalas*

Complexity Measure
keith
wnbd4

Inter-Cultural

Table 6.13 contains those measures which appear in more than one outlier pair for the inter-cultural comparisons made with the African *Timelines* to both the Indian *Decitalas* and the North Indian *Talas*.

Table 6.13: *Summary table of complexity measures which are listed as outliers in more than one pair from the boxplots of the comparison of African* Timelines *with 12 or 16 pulses to: (a) the Indian* Decitalas *with 12 or 16 pulses, (b) all of the Indian* Decitalas*, and (c) the North Indian* Talas*.*

(a) Indian *Decitalas* with 12 or 16 pulses.

Complexity Measure
hrun
ioi-g-mm
metricalponorm
oddity

(b) Indian *Decitalas*

Complexity Measure
offbeatness

(c) North Indian *Talas*

Complexity Measure
metricalonorm
wnbd

6.3 Random

The third part of the results applies the same method as discussed in § 6.2; however, here *random rhythms* are used to test for robustness among the complexity measures. In the first subsection, we investigate the stability of the complexity measures between random rhythms with 12 pulses, compared to random rhythms with 16 pulses. The second subsection performs a comparison of the combined data set of all the random rhythms to the three cultural data sets: (1) African *Timelines*, (2) Indian *Decitalas*, and (3) North Indian

Talas. The aim of the results in this part is to compare the performance of the complexity measures according to randomly generated rhythms and rhythms from different cultures.

6.3.1 Random vs. Random

Here we compare rhythms of 12 pulses with those of 16 pulses. For each data set, we calculated the complexity with a handful of measures. The phylogenetic trees are shown in Figure 6.17. Figure 6.17(a) shows the complexity measures on the random rhythms with 12 pulses, and Figure 6.17(b) shows the complexity measures on the random rhythms with 16 pulses.

Figure 6.17, presents general similarities in the clustering, along with differences. Some of the differences are more dramatic than others, such as the clustering of *dswap4*, which dramatically switches from a cluster with *dswap2* in Figure 6.17(a), to a cluster with *dswap* in Figure 6.17(b). However, a more subtle difference among the trees is *pressing*, which switches from a cluster with *keith* in Figure 6.17(a), to a cluster with *wnbd2* in Figure 6.17(b). However, the position of the cluster of *pressing* and *wnbd2* is still quite close to *keith*. In order to quantify what constitutes a significant change in position in the tree, we use a boxplot of the change in coefficients, $|\Delta_{rs}|$, to discover outliers. Again, these outliers are hypothesized to not perform robustly (i.e., are unstable) between data sets. Figure 6.18 shows the boxplot, which contains outliers past the cutoff of $|\Delta_{rs}| > 0.509$. We list the pairs which significantly change cluster positions between the trees, which we label as outliers in the data, in Table 6.14.

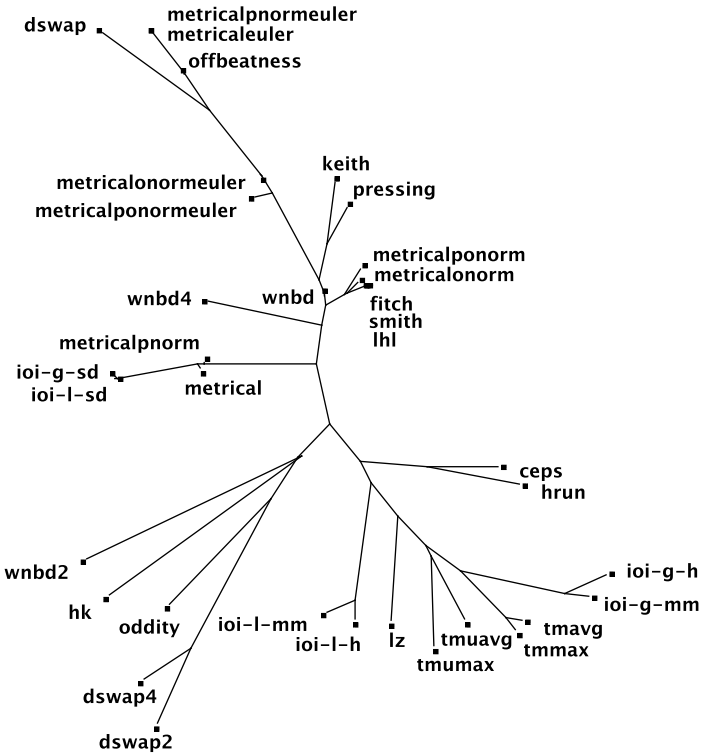
Table 6.14 lists the measures which are considered outlier pairs. Frequent among the pairs is the *wnbd2* measure, as well as the *ioi-l-h* measure. Visually, we can see that in Figure 6.17(a) *wnbd2* does not cluster with any measure, and is in a branch far from the other measures. However, in Figure 6.17(b), this is not the case; *wnbd2* moves quite close to *keith*.

6.3.2 Random vs. Cultural

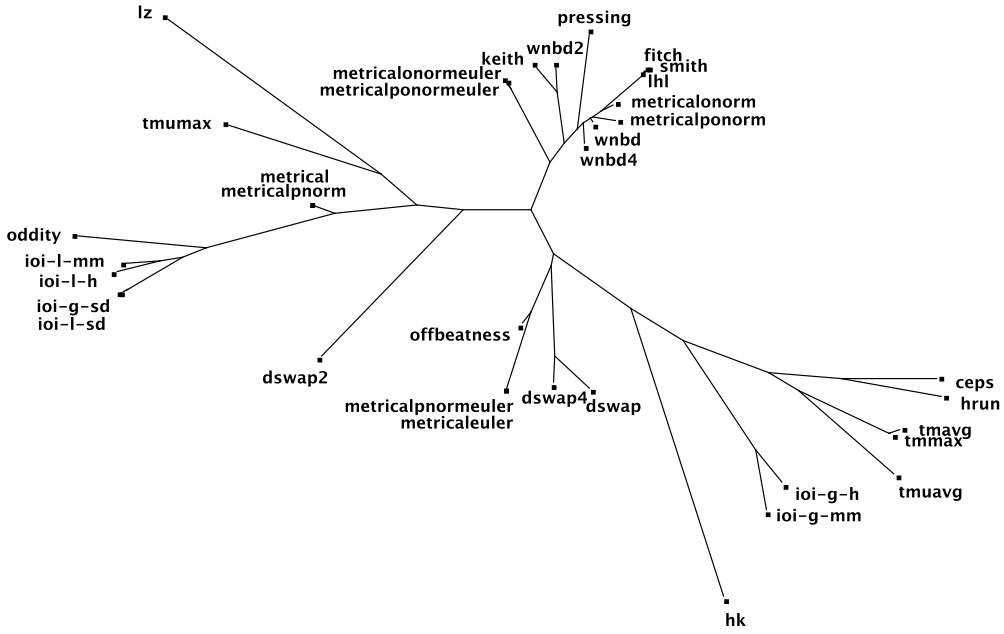
Using the randomly generated rhythms, we take the combined set of random rhythms, and compare this set to the African *Timelines*, the Indian *Decitalas*, and the North Indian *Talas*. The goal of this subsection is to test the complexity measures for robustness (i.e., stability) across data sets of rhythms from different cultures. Below we present the phylogenetic trees for the random rhythms and for each cultural data set. Moreover, we show boxplots of each comparison and list the outliers, which indicate unstable performance.

African *Timelines*

The first test of the random rhythms to a cultural data set, is seen here with the African *Timelines*. The African *Timeline* data set containing rhythms of 12 and 16 pulses is used. Consider the two phylogenetic trees in Figure 6.19 which shows the tree for the random



(a) BIONJ phylogenetic tree of the complexity measures on random rhythms with 12 pulses, least-squares fit 92.8.



(b) BIONJ phylogenetic tree of the complexity measures on random rhythms with 16 pulses, least-squares fit 94.2.

Figure 6.17: *BIONJ phylogenetic trees of the complexity measures which may be used on random rhythms of 12 and 16 pulses.*

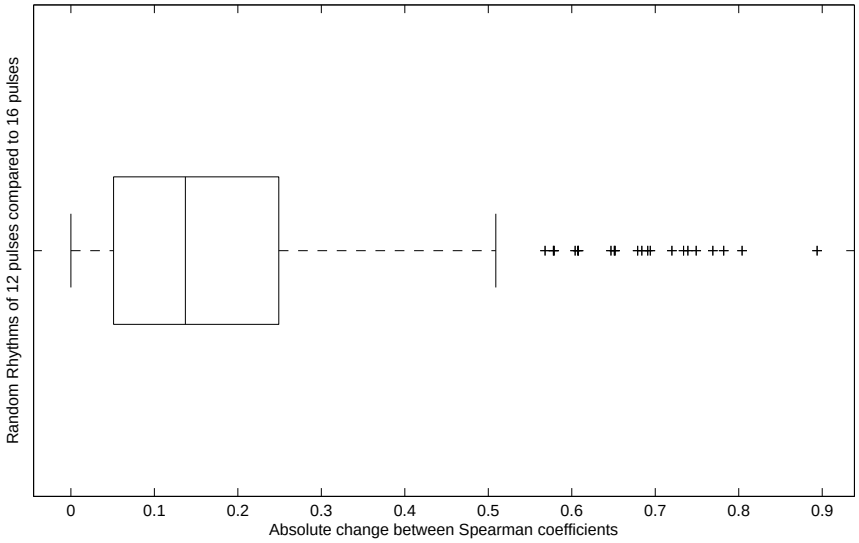


Figure 6.18: *Boxplot of the absolute value of the change ($|\Delta_{rs}|$) between corresponding pairs of Spearman rank correlation coefficients of complexity measures between random rhythms with 12 and 16 pulses, outlier cutoff $|\Delta_{rs}| > 0.509$.*

Table 6.14: *Outliers (as indicated by the boxplot in Figure 6.14) with an absolute-value change, $|\Delta_{rs}| > 0.509$, in Spearman rank correlation coefficient going from the random rhythms with 12 pulses (Figure 6.17(a)) to those with 16 pulses (Figure 6.17(b)).*

Complexity Measure Pair		$\lvert\Delta_{rs}\rvert$
wnbd2	keith	0.894
wnbd2	wnbd	0.804
wnbd2	metricalponorm	0.782
dswap4	pressing	0.769
wnbd2	metricalonorm	0.749
wnbd2	offbeatness	0.739
wnbd2	smith	0.734
dswap	dswap2	0.720
wnbd2	lhl	0.694
wnbd2	fitch	0.691
wnbd2	metricalponormeuler	0.684
dswap4	dswap	0.679
wnbd2	pressing	0.652
wnbd2	metricalonormeuler	0.651
tmumax	ceps	0.647
ioi-l-h	ioi-g-h	0.608
dswap4	wnbd	0.607
dswap4	metricalonorm	0.604
dswap4	metricalonormeuler	0.579
dswap4	metricalponorm	0.578
ioi-l-h	ioi-g-mm	0.568

rhythms in Figure 6.19(a), and then shown again for convenience in visual comparison, is the tree for the African *Timelines* in Figure 6.19(b).

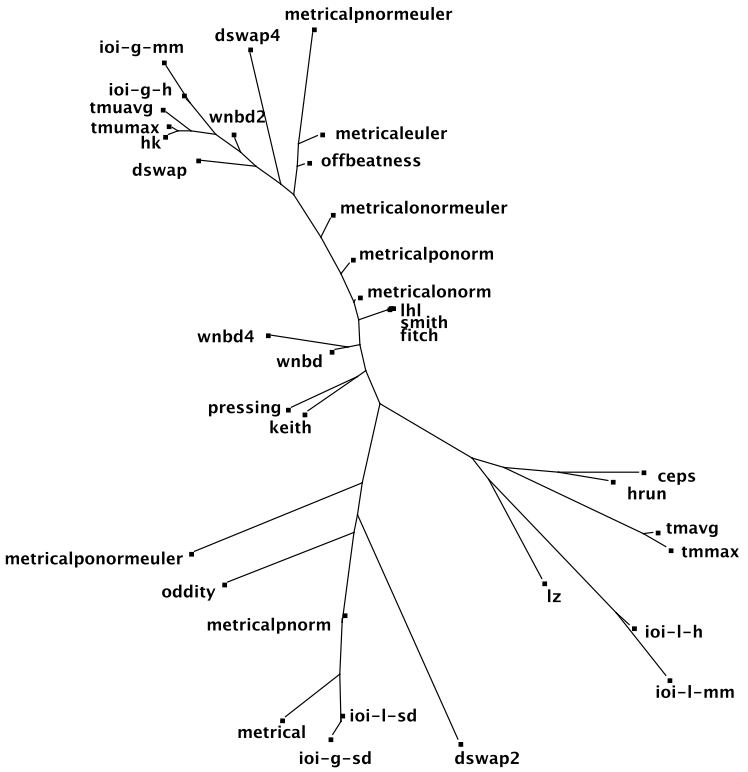
Between the phylogenetic trees in Figure 6.19, there are similarities among how the measures are clustered. For example, we can see that measures *fitch, keith, lhl, pressing, smith*, and *wnbd* cluster towards the middle of the tree in Figure 6.19(a). Those same measures also cluster toward the middle of the tree (with *metricalonorm* moving a bit closer) in Figure 6.19(b). We can discover similarities like the one just discussed by inspecting the tree, but we are interested in finding differences which are significant between the two trees. Generating a boxplot reveals any outliers which have this significant change. Figure 6.20 depicts this boxplot.

From the boxplot in Figure 6.20, the cutoff for outliers is $|\Delta_{rs}| > 0.417$. Moreover, we can see the outliers marked with a plus symbol. Consider Table 6.15 which lists the outliers found by this method.

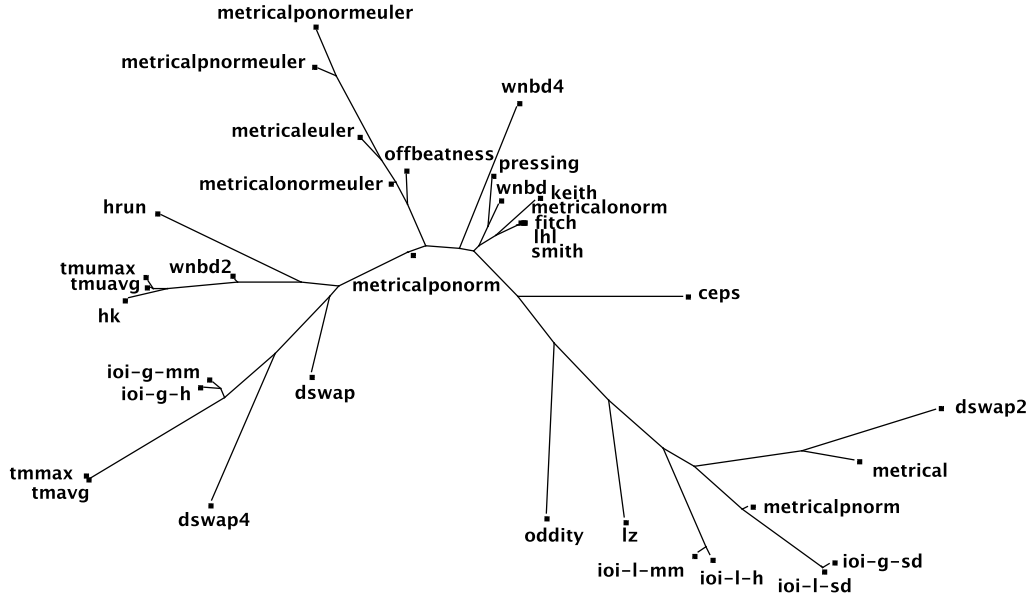
Table 6.15: *Outliers (as indicated by the boxplot in Figure 6.20) with an absolute-value change, $|\Delta_{rs}| > 0.417$, in Spearman rank correlation coefficient going from the random rhythms with 12 and 16 pulses (Figure 6.19(a)) to African* Timelines *with 12 and 16 pulses (Figure 6.19(b)).*

Complexity Measure Pair		$\lvert\Delta_{rs}\rvert$
hrun	ioi-l-mm	0.840
hrun	ioi-l-h	0.774
tmavg	oddity	0.592
hk	ioi-g-h	0.540
hrun	ceps	0.503
pressing	oddity	0.490
tmuavg	ioi-g-sd	0.485
lz	ioi-l-sd	0.476
hrun	ioi-l-sd	0.475
lz	hrun	0.470
lz	ioi-g-sd	0.468
tmumax	ioi-g-sd	0.463
tmmax	oddity	0.460
hk	offbeatness	0.449
lz	ioi-g-mm	0.439
hk	ioi-g-mm	0.420
lz	ioi-g-h	0.418

From the list of outlier pairs in Table 6.15, we can see that measures *hk, hrun, ioi-g-h, ioi-g-mm, ioi-g-sd, ioi-l-sd, lz*, and *oddity* are the most frequent in the outlier pairs. Taking *hrun* for example, in Figure 6.19(b), the measure clusters quite close to *ceps*, and is in a larger cluster with *ceps, ioi-l-h, ioi-l-mm, lz, tmavg*, and *tmmax*. However, Figure 6.19(b), shows *hrun* quite far from *ceps* and in a larger cluster of *hk, tmavg, tmmax*, and *wnbd2*.



(a) BIONJ phylogenetic tree of the complexity measures on random rhythms, least-squares fit 90.4.



(b) BIONJ phylogenetic tree of the complexity measures on African *Timelines* with 12 and 16 pulses, least-squares fit 94.2.

Figure 6.19: *BIONJ phylogenetic trees of the complexity measures which may be used on random rhythms of 12 and 16 pulses and African* Timelines *with 12 and 16 pulses.*

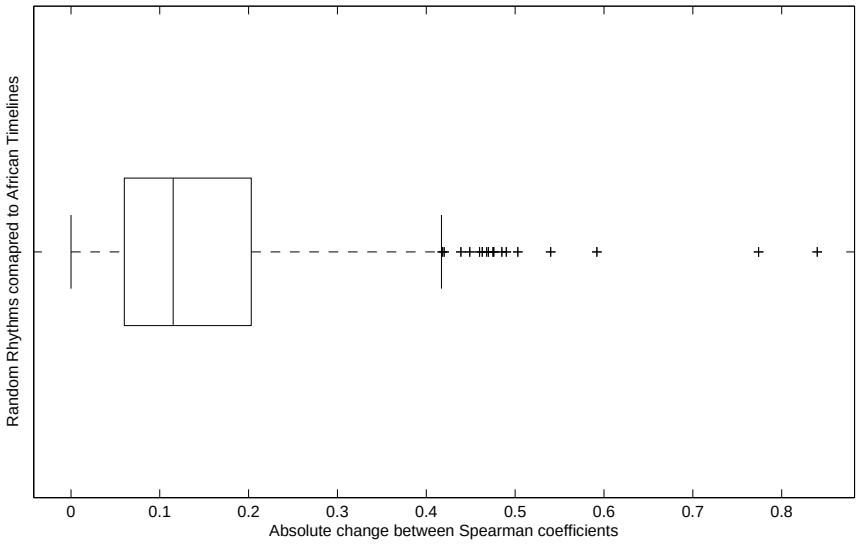


Figure 6.20: *Boxplot of the absolute value of the change ($|\Delta_{rs}|$) between corresponding pairs of Spearman rank correlation coefficients of complexity measures of random rhythms and African* Timelines *with 12 and 16 pulses, outlier cutoff* $|\Delta_{rs}| > 0.417$.

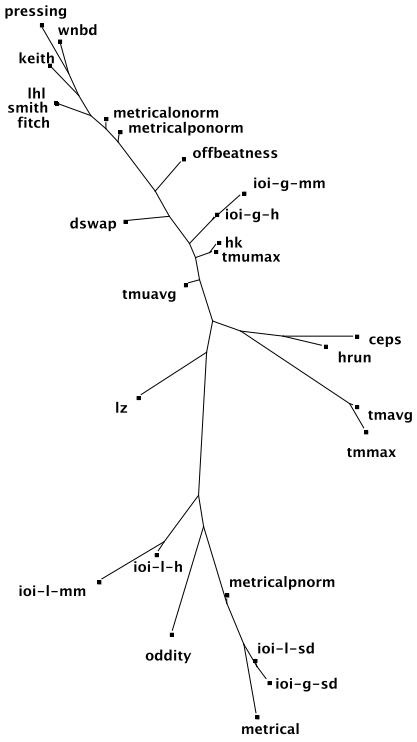
Indian *Decitalas*

We now turn to comparing the random rhythms to the Indian *Decitalas*. Figure 6.21 depicts the phylogenetic trees of the random rhythms in Figure 6.21(a), and also the Indian *Decitalas* in Figure 6.21(b). Note that both are repeat for convenience in visually comparing the trees.

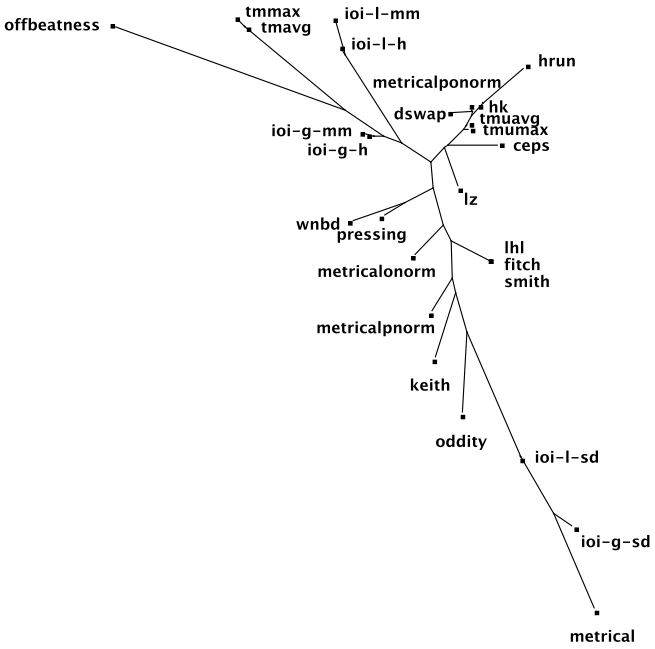
Inspecting the phylogenetic trees in Figure 6.21 shows similarities in clustering, and also differences. However, perhaps the most noticeable difference is with the *offbeatness* measure which is tucked into the upper part of the tree in Figure 6.21(a), but is quite distant residing at the top of the tree in Figure 6.21(b). However, even though this appears to be a significant deviation, we quantify this change with a boxplot of all the significant changes in coefficients between the two data sets. Consider this boxplot shown in Figure 6.22.

From the boxplot in Figure 6.22, we can see that the cutoff for outliers is $|\Delta_{rs}| > 0.695$. Moreover, the plus symbols indicates there are significant outliers in the data. In order to find out if the *offbeatness* measure did in fact have a significant change corresponding to the visual deviation in the phylogenetic trees, consider Table 6.16 which lists those pairs of measures with a significant deviation (i.e., the outliers) among all the $|\Delta_{rs}|$ values.

Table 6.16 in fact shows that the *offbeatness* measure is an outlier and appears in most of the outlier pairs. Thus we have a clear outlier in terms of change in coefficient between the random rhythms and the Indian *Decitalas*.



(a) BIONJ phylogenetic tree of the measures on random rhythms, least-squares fit 87.3.



(b) BIONJ phylogenetic tree of the complexity measures on Indian *Decitalas*, least-squares fit 95.5.

Figure 6.21: *BIONJ phylogenetic trees of the complexity measures which may be used on random rhythms with 12 and 16 pulses and Indian* Decitalas*.*

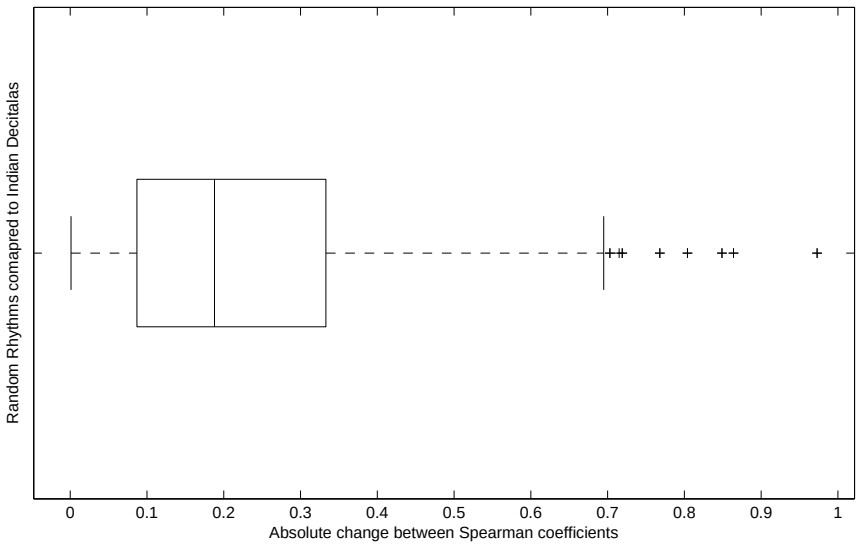


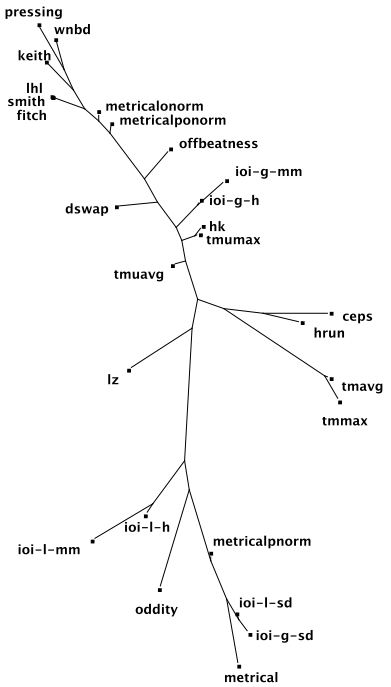
Figure 6.22: *Boxplot of the absolute value of the change ($|\Delta_{rs}|$) between corresponding pairs of Spearman rank correlation coefficients of complexity measures of random rhythms with 12 and 16 pulses and Indian* Decitalas*, outlier cutoff* $|\Delta_{rs}| > 0.695$*.*

Table 6.16: *Outliers (as indicated by the boxplot in Figure 6.22) with an absolute-value change,* $|\Delta_{rs}| > 0.695$*, in Spearman rank correlation coefficient going from the random rhythms with 12 and 16 pulses (Figure 6.21(a)) to all of the Indian* Decitalas *(Figure 6.21(b)).*

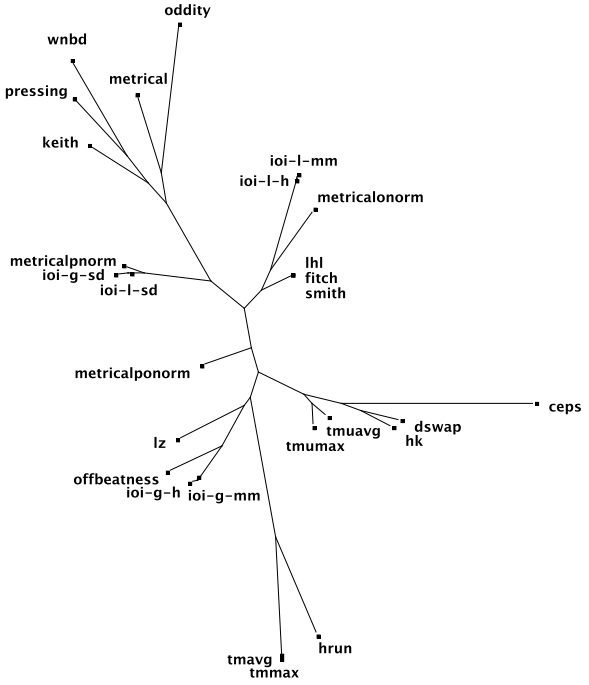
Complexity Measure Pair		$\lvert\Delta_{rs}\rvert$
offbeatness	hk	0.973
offbeatness	metricalponorm	0.864
offbeatness	dswap	0.849
offbeatness	tmumax	0.804
offbeatness	hrun	0.768
offbeatness	tmuavg	0.719
metricalpnorm	metrical	0.715
offbeatness	fitch	0.703

North Indian *Talas*

The North Indian *Talas* are now used as the data set of cultural rhythms for comparison to the random rhythms. Figure 6.23 shows the phylogenetic tree of the complexity measures on the random rhythm data set in Figure 6.23(a), along with the tree of the North Indian *Talas* data set in Figure 6.23(b). These figures have been previously shown; however, they are placed together here in order to compare the similarities and differences in the trees.



(a) BIONJ phylogenetic tree of the measures on random rhythms, least-squares fit 87.3.



(b) BIONJ phylogenetic tree of the complexity measures on North Indian *Talas*, least-squares fit 93.2.

Figure 6.23: *BIONJ phylogenetic trees of the complexity measures which may be used on random rhythms with 12 and 16 pulses and North Indian* Talas*.*

From Figure 6.23, we can see that there are a few notable differences between the trees. Take for example the *wnbd* measure. In Figure 6.23(a), we see this measure towards the top of the tree clustered near *fitch, keith, lhl*, *pressing*, and *smith*. However, in Figure 6.23(b), we see the measure at the top of the tree again clustered with *keith, pressing*, but quite far from *fitch, lhl*, and *smith*. Consider the boxplot in Figure 6.24 which show outliers in the data.

The boxplot in Figure 6.24 shows that there are outliers in the distribution of change in coefficient between the random rhythms and the North Indian *Talas*. Consider Table 6.17 which shows these outlier pairs.

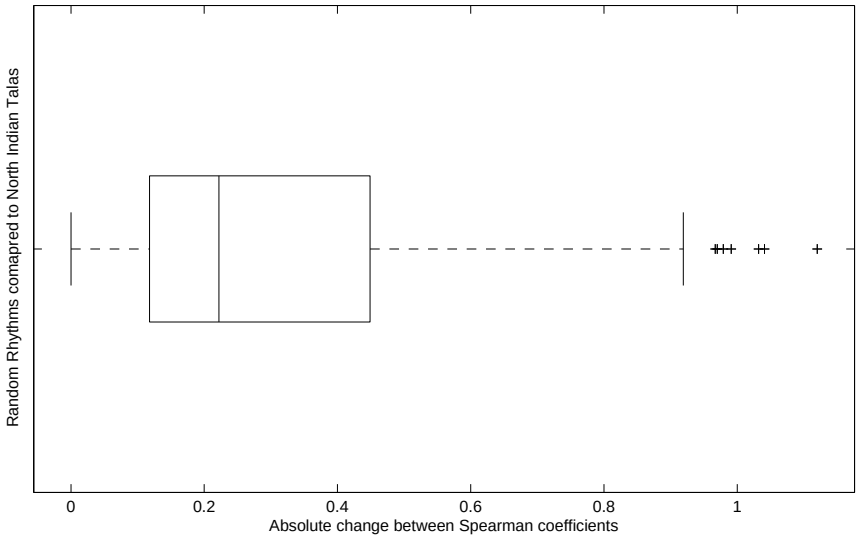


Figure 6.24: *Boxplot of the absolute value of the change ($|\Delta_{rs}|$) between corresponding pairs of Spearman rank correlation coefficients of complexity measures of random rhythms with 12 and 16 pulses and North Indian* Talas*, outlier cutoff* $|\Delta_{rs}| > 0.919$*.*

Table 6.17: *Outliers (as indicated by the boxplot in Figure 6.24) with an absolute-value change,* $|\Delta_{rs}| > 0.919$*, in Spearman rank correlation coefficient going from the random rhythms with 12 and 16 pulses (Figure 6.23(a)) to all of the North Indian* Talas *(Figure 6.23(b)).*

Complexity Measure Pair		$\lvert\Delta_{rs}\rvert$
keith	metricalonorm	1.120
hrun	ceps	1.041
pressing	metricalonorm	1.032
wnbd	metricalonorm	0.991
wnbd	lhl	0.979
wnbd	fitch	0.970
wnbd	smith	0.967

In Table 6.17, we can see that *wnbd* is in fact an outlier and is frequent in the pairs. Also, this points to the *metricalonorm* which has significant coefficient change with three other measures.

6.3.3 Summary

We observed the relationships between the complexity measures in two ways. First we looked at the relationship of the measures on random rhythms where the number of pulses changed. This was shown for random rhythms with 12 and 16 pulses. Second we looked at the relationship of the complexity measures on these random rhythms compared to different cultures. We made three comparisons of the random rhythms: (1) African *Timelines*, (2) Indian *Decitalas*, and (3) the North Indian *Talas*. Below we provide a table that summarizes which complexity measures are in a pair of at least two of the outliers in the boxplots, and outlier tables for each respective comparison.

Random vs. Random

Table 6.18 below consists of those complexity measures which are considered to be in an outlier pair for the robustness test within the random rhythm data.

Table 6.18: *Summary table of measures listed as outliers in more than one pair from the boxplot of the random rhythms with 12 pulses compared to 16 pulses.*

Complexity Measure
dswap4
ioi-l-h
wnbd2

Random vs. Cultural

Table 6.19 contains those measures appearing in more than one outlier pair comparing the random rhythms to the African *Timelines*, Indian *Decitalas*, and North Indian *Talas*.

Table 6.19: *Summary table of complexity measures which are listed as outliers in more than one pair from the boxplots of the comparison of random rhythms with 12 or 16 pulses to the cultural data.*

(a) African *Timelines*

Complexity Measure
hk
hrun
ioi-g-h
ioi-g-mm
ioi-g-sd
ioi-l-sd
lz
oddity

(b) Indian *Decitalas*

Complexity Measure
offbeatness

(c) North Indian *Talas*

Complexity Measure
metricalonorm
wnbd

Chapter 7

Discussion

This chapter presents a discussion of the results previously reviewed. First, the human-based complexity measures are discussed and general conclusions are drawn. Second, the culturally-based comparisons are discussed and conclusions are made based on such results. Third, the random-based comparisons are reviewed and discussed, and conclusions are drawn. Finally, concluding remarks are made and open problems for future work are presented.

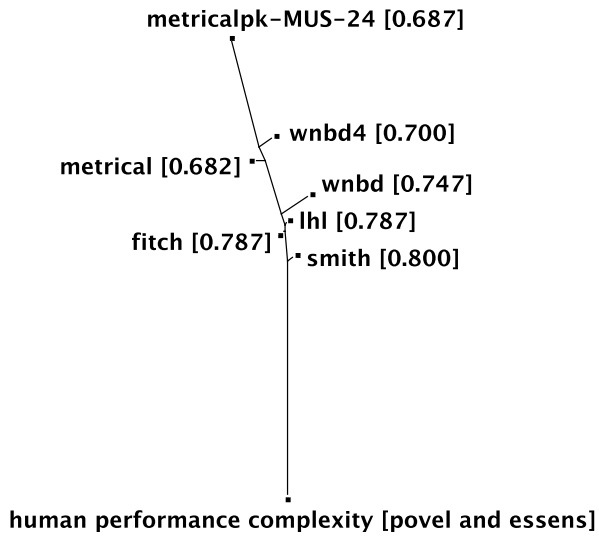
7.1 Human Complexity

The first goal of this work has been to validate the complexity measures presented against human-based measures of rhythm complexity. Computational measures were compared to measures of human-based performance complexity, human-based perceptual complexity, and human-based metrical complexity, as well as to each other. We discuss the results obtained regarding each human complexity measure from Chapter 6 and propose open problems for future work. Consider each discussion in the following subsections.

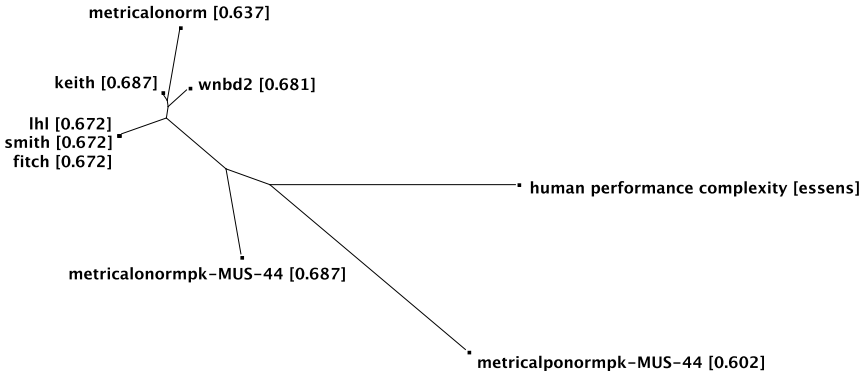
7.1.1 Performance

We have compared the complexity measures to human-based performance complexity from all three of the psychological data sets. From the comparison we have similarities and differences regarding which measures of complexity performed most closely to the human measures. Using the data from Povel and Essens, we have Smith and Honing's version of the Longuet-Higgins and Lee measure performing best ($r_s^* = 0.800$). For the Essens data set, Toussaint's Metrical Complexity (Onset Normalized) using Palmer and Krumhansl's musicians-4/4 weighting scheme performed best ($r_s^* = 0.687$). For Fitch and Rosenfeld's data set, we have our unconstrained version of Tanguiane's Complexity using the average across metrical levels, performing best ($r_s^* = 0.613$). To see this, consider Figure 7.1 which depicts phylogenetic trees of the top measures in Table 6.4 in Chapter 6.

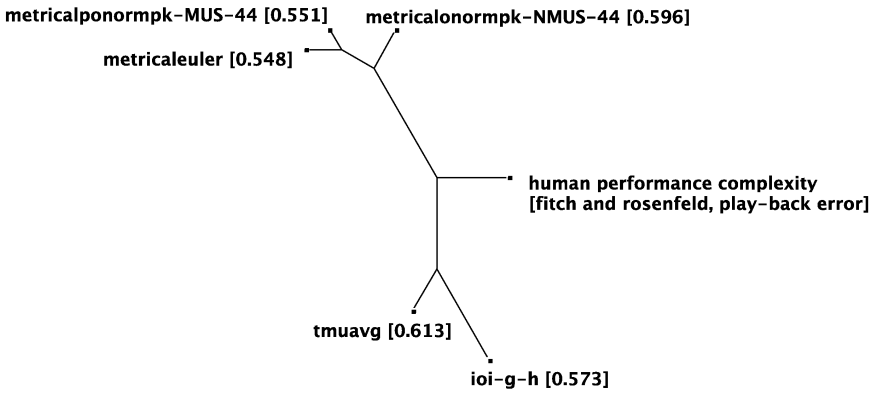
In Figure 7.1, it is clear that the human-based performance complexity in each tree does not cluster with any of the measures. This means that the intercorrelation between measures are higher than the correlation of the complexity measure to the human-based measure. First, look at the tree in Figure 7.1(a). Even though the correlation of *fitch,lhl*, and *smith* are all moderately high ($r_s^* > 0.750$), the measures still appear to have a large distance from Povel and Essens' human performance complexity. This is because the intercorrelations are quite high; i.e., $\bar{r}_s^* > 0.9$. Therefore, the measures are more closely clustered to each other than to the human-based measure.



(a) Povel and Essens, least-squares fit 99.0.



(b) Essens, least-squares fit 99.1.



(c) Fitch and Rosenfeld, least-squares fit 99.9.

Figure 7.1: *Phylogenetic trees using the BIONJ method of the complexity measures which obtained the highest performance indicated from Table 6.4 in Chapter 6. Here, we have the computational measures along with the human-based performance complexity from Povel and Essens in (a), Essens in (b), and Fitch and Rosenfeld in (c). Note that the coefficients in each tree are significant at the* $p \leq 0.001$ *level.*

Among the top-performing measures in Figure 7.1(a), all of them except *wnbd* and *wnbd4* use the metrical hierarchy. Also, Palmer and Krumhansl's weighting scheme of the musicians with a $2/4$ metrical structure has better performance than the $4/4$ weighting scheme. In addition, consider the data which was used to arrive at these results. The rhythms in the data set from Povel and Essens consisted of patterns which had: the same number of pulses, the same number of onsets, and the same distribution of interonset intervals. This suggests three possible conclusions:

1. Under controlled data human-based performance complexity may be best measured by algorithms which use the metrical hierarchy with the original weighting scheme.
2. Such restricted data (same IOIs, number of onsets and pulses) may not provide adequate data for the accurate validation of the complexity measures against the human-based performance complexity.
3. The complexity measure may be evaluating some other property that is related to human-based performance complexity, since the measures are so highly intercorrelated.

Let us now consider the phylogenetic tree in Figure 7.1(b). Here, the data set from Essens was used to compute the complexity using the measures shown in the tree. Each are the top performers listed in Table 6.4 from Chapter 6. Again the measures which occur in this tree are those which use the metrical hierarchy. However, *keith* and *wnbd2* stand out as those which are not based on this hierarchy. Note that *keith* is among the top-performers for the Essens data set, but not for the Povel and Essens data.

Taking into consideration the data from Essens, the rhythms all have the same number of pulses and have a high number of onsets (10.375 on average). Moreover, they were tested on a small number of participants in the study (6 participants [48]). In any case, the measure which performed most closely to the human-based measure is *metricalonormpk-MUS-44* with $r_s^* = 0.687$. What is noteworthy, is that even though the top performers are still those using the metrical hierarchy, the best performing measure is when the hierarchy uses Palmer and Krumhansl's musicians $4/4$ weighting scheme. Moreover, *keith*, which also has $r_s^* = 0.687$ is tied for top-performance, and does not rely on the metrical hierarchy weighting scheme. Though, the measure does use an arbitrary weighting scheme for musical events: 1 for *hesitation*, 2 for *anticipation*, and 3 for *syncopation*. So perhaps this is why the measure is pulled into the cluster of the measures using the metrical hierarchy with original weights; i.e., the cluster of *fitch, lhl, metricalonorm*, and *smith*. These results suggest the following conclusions.

1. The metrical hierarchy with Palmer and Krumhansl's musicians $4/4$ weights, most

accurately measures human-based performance complexity of rhythms with variation in the number of onsets.

2. The *fitch, keith, lhl, metricalonorm, smith*, and *wnbd* measures tightly cluster away from the human-based measure, because they may be measuring some other aspect of the rhythms, related to how difficult they are to perform.

Finally, examination of the phylogenetic tree in Figure 7.1(c) reveals a different story than the previous two trees. However, we do see a similarity in that the Metrical Complexity measures, which use the Palmer and Krumhansl weights, for both musicians and non-musicians in $4/4$, cluster together near the top of the tree. The main difference is that no measures using the original metrical weights are present among the top performers. In fact, two measures which previously obtained poor performance are seen here as top performers: *tmuavg* and *ioi-g-h*. Tanguiane's Complexity pertains to musical elaboration, whereas the IOI measure, pertains to the entropy of the IOI histogram. Thus, the results are split where those measures using an alternate weighting scheme for the metrical hierarchy are on the top of the tree, and those measures which have previously not performed well, are on the bottom of the tree. The human-based measure is in the middle of the split.

This dramatic change in which measures perform best is perhaps due to the data set provided by Fitch and Rosenfeld. Their data was generated to yield a set of rhythms which varied in the complexity according to Fitch and Rosenfeld's implementation of the Longuet-Higgins and Lee measure. By doing so, all the rhythms generated had the same number of pulses, a smaller number of onsets (4.8 on average), and in fact were more representative of rhythms from African and Indian cultures. Table 7.1 lists the rhythms found in African and Indian culture. Note that the first column is the rhythm ID number from Fitch and Rosenfeld's data in Table 4.3 in Chapter 4, the second column is the name of the rhythm for African (on top) or Indian (below) cultures. The third and fourth columns show the rhythm in the culture and the the rhythm from Fitch and Rosenfeld's data set.

As can be seen in Table 7.1, most of the rhythms from Fitch and Rosenfeld's data set are rotations of the rhythms found in African or Indian culture. Note that a rotation is defined as placing a rhythm on a circle, and then turning the circle by a rotational distance based on the pulse. Turning the circle clockwise by one rotation, for a rhythm with 16 pulses, would bring pulse 0 to pulse 1, pulse 1 to pulse 2, ..., and pulse 15 to pulse 0. However, not all the rhythms from Fitch and Rosenfeld have to be rotated, the precise rhythms of the *Rumba* from African culture and the *Pratyanga* from Indian culture are present in the set. Moreover, the duplicate names in the table, means that some rhythms from Fitch and Rosenfeld's data are rotations of each other. The results of the representative data, from Fitch and Rosenfeld's human performance complexity, suggests the following.

Table 7.1: *Rhythms from Fitch and Rosenfeld's data set which are found also in African or Indian cultures.*

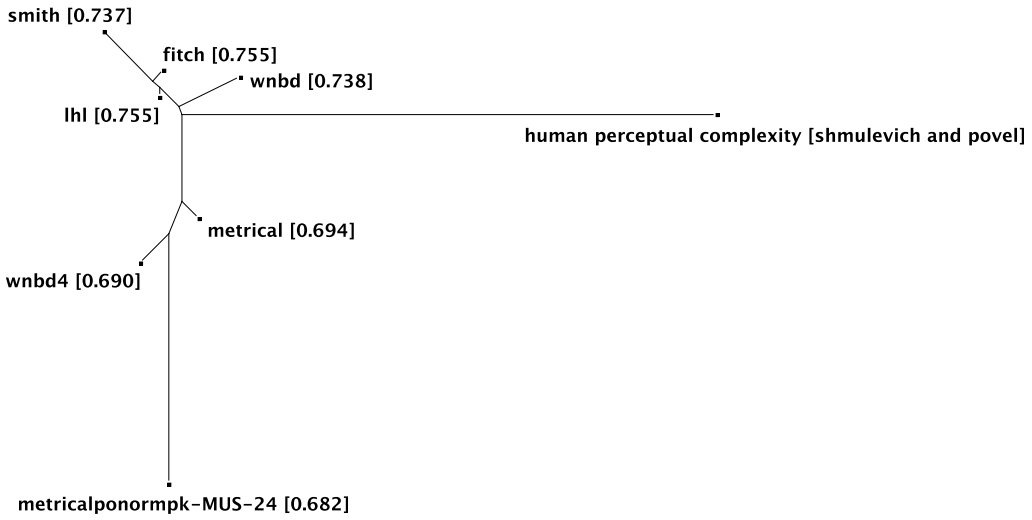
No.	Name	Rhythm	Fitch/Rosenfeld
12	Domba	x.x..x...x...x..	.x...x...x..x.x.
12	Kpatsa	x...x...x..x.x..	.x...x...x..x.x.
6	Kromanti	x.x...x...x.x...	..x...x.x...x.x.
12	Rap-Drum	x...x..x.x..x...	.x...x...x..x.x.
8	Rumba	x..x...x..x.x...	x..x...x..x.x...
6	Shiko	x...x.x...x.x...	..x...x.x...x.x.
6	Timini	x.x...x.x...x...	..x...x.x...x.x.
2	Tuareg	x...x.x.x...x...	x...x...x...x.x.
22	Tuareg	x...x.x.x...x...	..x...x.x.x...x.
30	Tuareg	x...x.x.x...x...	.x.x...x...x...x
6	Jayacri	x...x.x...x.x...	..x...x.x...x.x.
2	Pratyanga	x...x...x...x.x.	x...x...x...x.x.
22	Pratyanga	x...x...x...x.x.	..x...x.x.x...x.
30	Pratyanga	x...x...x...x.x.	.x.x...x...x...x
6	Simhanada	x.x...x...x.x...	..x...x.x...x.x.
2	Vijayananda	x.x.x...x...x...	x...x...x...x.x.
22	Vijayananda	x.x.x...x...x...	..x...x.x.x...x.
30	Vijayananda	x.x.x...x...x...	.x.x...x...x...x

1. Measures which use the metrical hierarchy and the human judged weighting scheme of musicians or non-musicians in $4/4$, perform better than the orignal weighting of the metrical hierarchy.
2. The global interonset histogram may provide valuable information in terms of measuring human-based performance complexity of rhythms.
3. The notion of musical elaboration presented by Tanguiane may be significant when measuring the complexity of rhythm performance.
4. None of the complexity measures adequately or even accurately measure human-based performance complexity when rhythms from African or Indian cultures are concerned, since the highest correlation to the human-based measure is roughly 0.6.

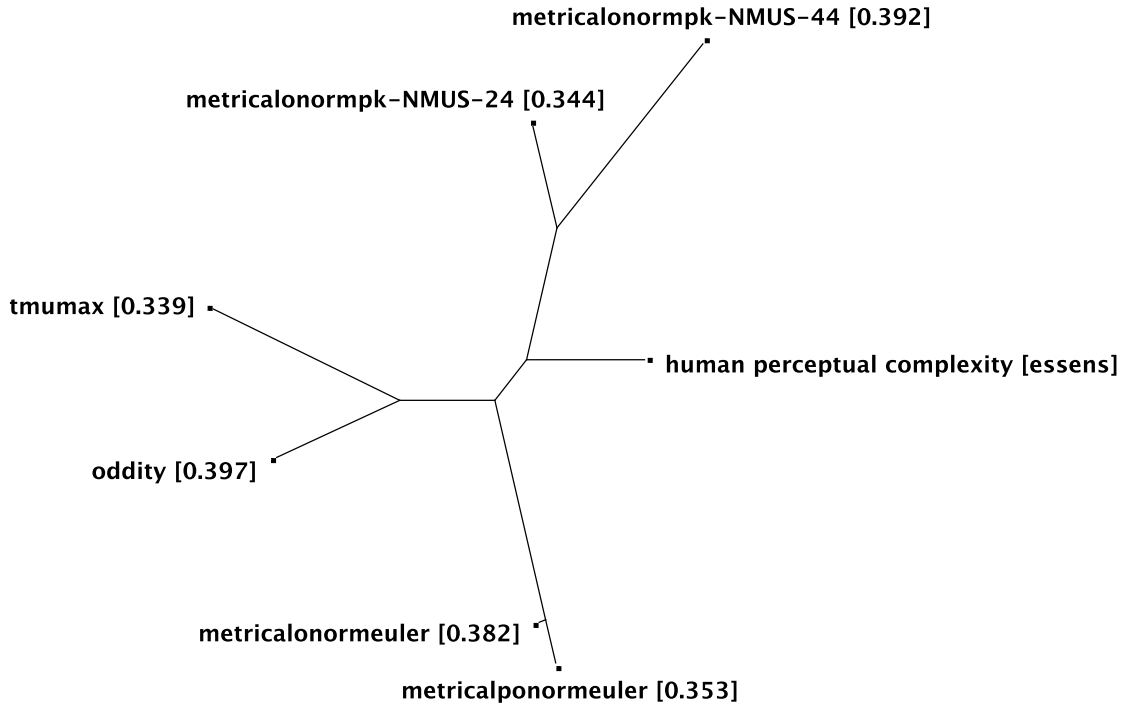
7.1.2 Perceptual

Human-based perceptual complexity was measured by two psychological studies, and thus far we have made comparisons of computational measures to both of these human measures. Figure 7.2 shows phylogenetic trees with respect to each data set: the Shumlevich and Povel data in Figure 7.2(a) and the Essens data in Figure 7.2(b). We show the top measures for each, which were listed in Table 6.5 in Chapter 6. For convenience, the Spearman rank correlation coefficient for each computational measure compared to human-based perceptual complexity is shown appended to each node label in square brackets.

First, consider Figure 7.2(a) which shows the best results from Shmulevich and Povel's human judgments of perceptual complexity. The measure that performs most closely to the



(a) Shmulevich and Povel, least-squares fit 98.9.



(b) Essens, least-squares fit 97.8.

Figure 7.2: *Phylogenetic trees using the BIONJ method of the complexity measures which obtained the highest performance indicated from Table 6.5 in Chapter 6. Here, we have the computational measures along with the human-based perceptual complexity from Shmulevich and Povel in (a) and Essens in (b). Note that the coefficients of the measures in (a) are significant at the* $p \leq 0.001$ *level, whereas in (b), only* oddity *has significance of* $p \leq 0.05$ *while the rest are not significant with* $p \leq 0.1$.

human measure is the Longuet-Higgins and Lee measure, including the implementations by Fitch and Rosenfeld [57], and Smith and Honing [150]. The other measures which perform well are those which are based on the metrical hierarchy; e.g., Toussaint's Metrical Complexity and the LHL measures. The WNBD measure, not based on the metrical hierarchy, also performs well.

However, the phylogenetic tree shows that such top performing measures are still set apart from the human-based perceptual complexity. The computational measures obtain a correlation of roughly $r_s^* \approx 0.7$, which indicates a moderate relationship to the human-based measure. However, the relationship among the computational complexity measures is much higher. Disregarding the human-based measure correlations, the intercorrelation of the measures is over $r_s^* > 0.9$. This indicates that the measures are similarly ranking the rhythms, where the ranking moderately correlates to the human judgments, but is perhaps measuring some other property of the patterns to determine their complexity, which may not accurately reflect human perceptual judgments of complexity.

In addition, the data may provide insight to the performance of the measures. For example, the data set of rhythms was very strictly generated. Each rhythm had the same number of pulses, onsets, and intervals; i.e., they were generated by permuting a single interval set [119]. Considering the data set and also the correlation results, the following conclusions are suggested.

1. Complexity measures which have the highest performance regarding human-based perceptual complexity of rhythms, use the metrical hierarchy.
2. When the complexity of constrained rhythms is measured, the original weights, from Lerdahl and Jackendoff, of the metrical hierarchy yield the most accurate results.

Second, consider Figure 7.2(b), which shows the best results regarding the human judgments of complexity from Essens' study [48]. The results here are poor in terms of correlation; however, the phylogenetic tree may still provide valuable information. Even though the correlation with human-based perceptual complexity is low, the intercorrelations are still of interest. Take for instance Toussaint's Metrical Complexity using the Palmer and Krumhansl non-musician weighting in $2/4$ and $4/4$, and using the Euler weighting scheme. Here we see both groups are split by the human-based perceptual complexity; thus they perform similarly to the human judgments, but are not similar to each other. This shows that using alternate weighting schemes can dramatically change the relationship to the other measures (put them on opposite sides of the tree), but using a standard framework (i.e., a metrical hierarchy) keeps the measures at the same distance to the human judgments. Also, two measures which generally do not perform well, *tmumax* and *oddity*, are seen as top performers regarding Essens' Human Perceptual Complexity.

There is a noted difference in the rhythm data used in Essens' study compared to the data used in Shmulevich and Povel's study. In Essens, the rhythms varied in the number of onsets. The mean number of onsets is 10.375 and the standard deviation of 1.056. As a result of this data set, dropping the constraints imposed by the nature of Povel and Essens' data, may allow for a more realistic (in terms of rhythms found among cultures) sense of performance. Thus, we may conclude the following.

1. Perceptual complexity judgments on rhythms which vary in the number of onsets yields results close to those computational measures which use an alternate weighting scheme for the metrical hierarchy, such as Palmer and Krumhansl's or Euler's scheme.
2. Properties of the Rhythmic Oddity measure or of Tanguiane's Complexity measure can perhaps be used for measuring human-based perceptual complexity, as they are the best (albeit still low) performers.

7.1.3 Metrical

The study by Fitch and Rosenfeld [57], also presents a new form of complexity which previous studies [48, 119, 146] did not address; i.e., human metrical complexity, which is the difficulty people have understanding the metrical structure of a rhythm. Figure 7.3 shows a phylogenetic tree of the top performing computational complexity measures compared to two human-based measures: Fitch and Rosenfeld's Human Metrical Complexity (Beat-Tapping Error Adjusted) and Fitch and Rosenfeld's Human Metrical Complexity (Number of Resets). Note that there are two coefficients in brackets of each complexity measure node label. The first is the coefficient compared to the Beat-Tapping Error Adjusted and the second is the coefficient compared to the Number of Resets. Also, the nodes in the tree represent the measures from Table 6.6.

Notice that Toussaint's Metrical Complexity dominates the top performing measures regarding both of the human-based metrical complexity measures. Also, not only do different versions of the Metrical Complexity measure dominate, but they also attain high correlation to the human measures, e.g., *metricalponorm* has $r_s^* = 0.834$ compared to the Beat-Tapping Error Adjusted and $r_s^* = 0.793$ compared to the Number of Resets. Moreover, as described in §§ 7.1.1, we have that some of the rhythms in Fitch and Rosenfeld's data set, represent rhythms found in both African and Indian cultures. From this we may conclude the following.

1. Complexity measures based on the metrical hierarchy with the original weighting scheme perform *accurately* to human-based metrical complexity on rhythms which contain a subset of rhythms from African and Indian cultures.

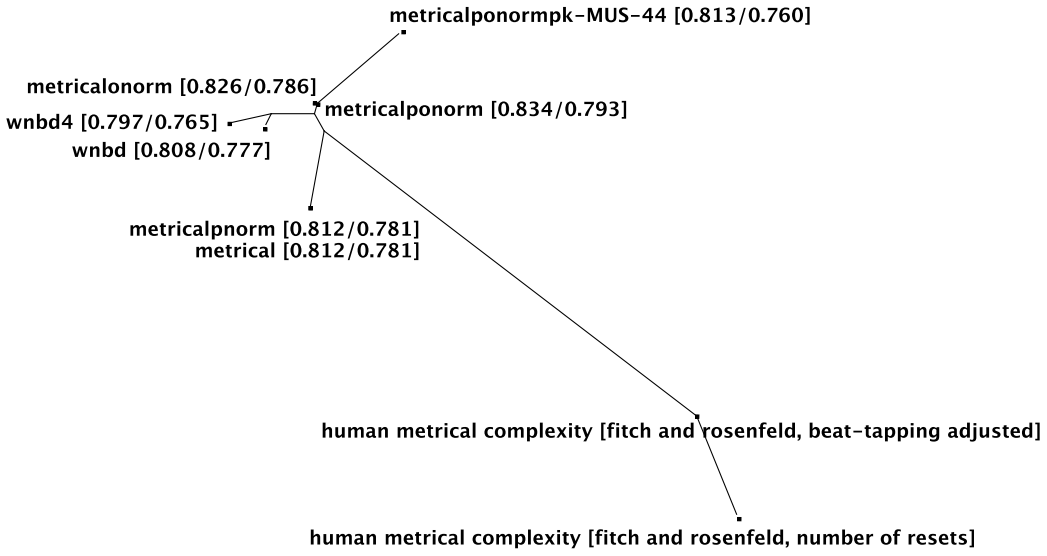


Figure 7.3: *Phylogenetic tree using the BIONJ method (least-squares fit 99.8) for the complexity measures which obtained the highest performance indicated from Table 6.6 in Chapter 6. Here, we have the computational measures along with the Human Metrical Complexity from Fitch and Rosenfeld regarding Beat-Tapping Error (Adjusted) and the Number of Resets. Note that the coefficients of the measures in both trees are significant at the* $p \leq 0.001$ *level.*

2. The $4/4$ metrical structure yields more accurate coefficients than $2/4$, as seen by *metricalponormpk-MUS-44* and *wnbd4*.
3. Human Metrical Complexity is the form of complexity closest to the computational complexity measures, 55% of the measures obtained a correlation $r_s^* > 0.5$ compared to the measure Beat-Tapping Error Adjusted.

7.2 Cultural Complexity

The second goal of this thesis was to validate the computational complexity measures in a cultural context. Thus we used representative data sets from African and Indian cultures. However, because we do not currently have human-based measures for any aspects of complexity for such rhythms, we chose to observe the behavior of the complexity measures between the different data sets. For instance, the robustness (i.e., stability) of the complexity measures was evaluated. Thus, the African *Timelines* were compared to both the Indian *Decitalas* and Indian *Talas*. In the following we discuss the results of such comparisons.

In Chapter 6, we evaluated the complexity measures' performance with respect to each other, between two data sets from within the same culture, and across cultures. In this evaluation we measured the difference in the intercorrelation coefficients going from one data set to the other, and used the basic technique of a boxplot to look for outlier pairs among the differences. Essentially, complexity measures which showed signs of unstable

performance were uncovered as outliers. This means that the correlation coefficient of a pair of measures in one data set was dramatically different compared to the coefficient of that same pair in the second data set. We will first discuss the results from the intra-cultural comparison, and second discuss the inter-cultural comparison.

7.2.1 Intra-Cultural Robustness

Consider Table 7.2 below which lists those measures which were frequent among the outlier pairs comparing sets of rhythms from the same culture. This list contains those measures that may be suspect of not having robust performance. Table 7.2 shows an outlier pair where the measure on the left may be paired with each measure on the right in the comma separated list. Note that each comparison is separated by a horizontal line.

Table 7.2: *Complexity measures which are considered unstable comparing the African* Timelines *with 12 pulses to those with 16 pulses, and below the horizontal separator are those unstable measures for the Indian* Decitalas *with 12 pulses compared to those with 16 pulses.*

Data Set	Complexity Measure Pair	
African *Timelines*	dswap2	dswap4, ioi-l-sd
	dswap4	dswap4, hrun
	keith	wnbd2, wnbd4
	wnbd2	fitch, lhl, keith, metricalonorm, metricalponorm, smith
Indian *Decitalas*	keith	fitch, lhl, smith, wnbd4, wnbd
	wnbd4	keith, tmuavg

One can immediately notice that *keith* is present as an unstable measure in both intra-cultural comparison, as is the *wnbd* measure, except with differing beat parameters (2 and 4). The reason for the the unstable performance of *keith* between data sets of 12 and 16 pulses is most likely due to the definition of the complexity measure. Recall that for the duration of each onset, the closest *power of two* is found which is smaller than the duration [86]. Using this scheme involving powers of two to determine the beat is problematic for rhythms with 12 pulses. This is because the beat does not typically fall on all powers of two as with rhythms of 16 pulses. Consider Figure 7.4 which shows the only possible metrical hierarchy for 16 pulses. This is the beat structure imposed by Keith's measure.

Figure 7.4 shows that the beat falls on a power of two at each level in the tree. Compare this now to the cases of where the beat may fall when a rhythm has 12 pulses. Figure 7.5 shows three different possible metrical hierarchies.

Notice in Figure 7.5 that the beat does not always fall on a power of two, sometimes it falls at a multiple of three. Hence, we see why Keith's measure may not be robust comparing rhythms with 12 pulses to those with 16; by Keith's definition, the measure does not always find the correct position of the beat.

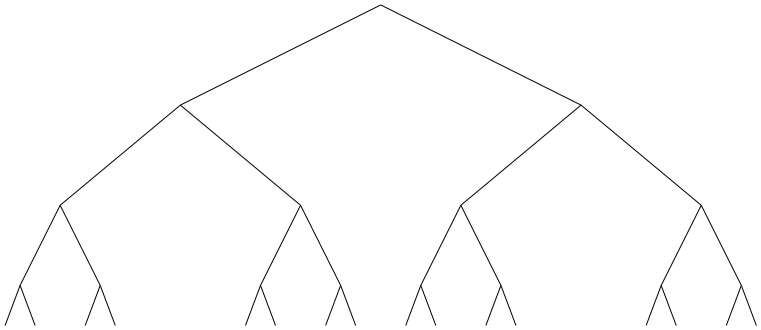
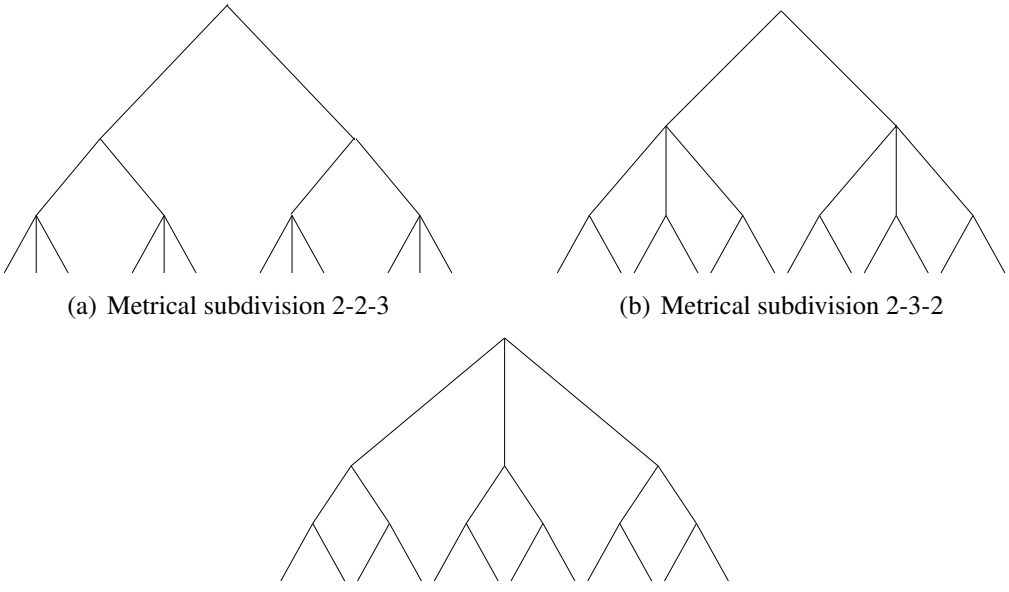


Figure 7.4: *Metrical hierarchy for the subdivision of 16 pulses.*



(a) Metrical subdivision 2-2-3

(b) Metrical subdivision 2-3-2

(c) Metrical subdivision 3-2-2

Figure 7.5: *Metrical hierarchies for each subdivision of 12 pulses.*

The beat position may also be the reason why *wnbd2* and *wnbd4* also show signs of instability between rhythms of 12 and 16 pulse. The *wnbd2* measure considers every second pulse to be the beat, whereas the *wnbd4* considers every fourth pulse to be the beat. Therefore, when one has a rhythm with 12 pulses, the beat may not fall at every two or every fourth pulse, especially as seen in the case presented in Figure 7.5(a) where the beat at the lowest metrical level can be see every three pulses. Similarly this is the most likely case for the unstable performance of the *dswap* measures as well. These results suggest the following.

1. The measure proposed by Keith [86] is not appropriate for rhythms of 12 pulses; however, if the definition were generalized, then perhaps it could still be applicable.
2. When comparing rhythms with 12 pulses to 16 pulses, measures which are parameterized with beat-placement, like *wnbd* and *dswap* should perhaps be giving a priori knowledge of which beat-placement parameters are appropriate for the given rhythm; such as three, in the case of 12 pulses and four, in the case of 16 pulses.

7.2.2 Inter-Cultural Robustness

Here we turn to a discussion of unstable rhythms between different cultures. The results show the African *Timelines* with 12 or 16 pulses compared to the Indian *Decitalas* with 12 or 16 pulses, the Indian *Decitalas*, and the North Indian *Talas*. Consider Table 7.3 which lists the unstable measures for each comparison.

Table 7.3: *Complexity measures which are considered unstable comparing the African* Timelines *with 12 or 16 pulses to Indian* Decitalas *with 12 or 16 pulses, all the Indian* Decitalas *below the first horizontal separator, and the North Indian* Talas *below the second horizontal separator.*

Data Set	Complexity Measure Pair	
Indian (12,16)-*Decitalas*	hrun	metricalonorm, wnbd
	ioi-g-mm	tmavg, tmmax
	metricalponorm	fitch, lhl, smith
	oddity	tmavg, tmmax
Indian *Decitalas*	offbeatness	fitch, hrun, keith, lhl, metricalponorm, smith
North Indian *Talas*	metricalonorm	keith, pressing, wnbd
	wnbd	fitch, lhl, metricalonorm, smith

Table 7.3 shows the pairs of complexity measures which were outliers going from one data set to the other. The complexity measure on the left paired with each measure on the right forms an outlier pair. Thus, for each pair we have that one measure or both have unstable performance across data sets. The table also divides the three comparisons by horizontal divider lines. The first four rows consist of those outlier pairs for the African

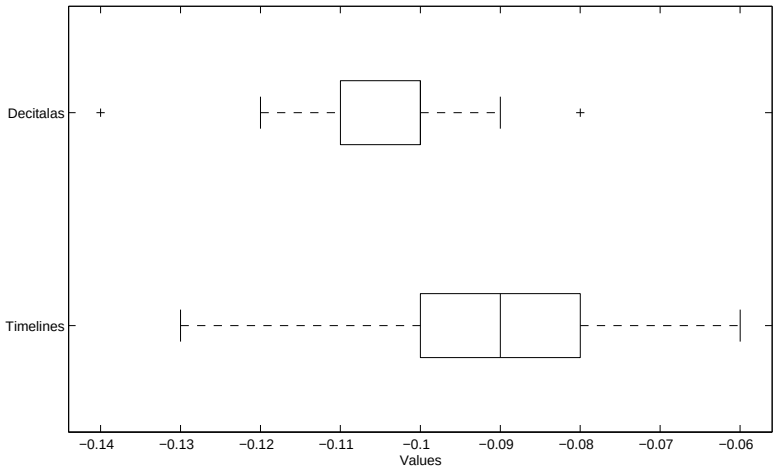
Timelines with 12 or 16 pulses compared to the Indian *Decitalas* with 12 or 16 pulses, the row beneath the first separator consists of those pairs of the African *Timelines* of 12 or 16 pulses compared to the Indian *Decitalas*, and the last two rows beneath the second separator consists of those outlier pairs of the comparison of the African *Timelines* of 12 or 16 pulses and the North Indian *Talas*.

First, consider the set of outlier pairs for the African *Timelines* compared to the Indian *Decitalas* of 12 or 16 pulses. For each pair, we can notice in the phylogenetic trees in Figure 6.11 from Chapter 6, in one tree the pairs are relatively close, whereas in the second tree, they are far apart. For example, take the pairs: (*metricalponorm, fitch*), (*metricalponorm, lhl*), and (*metricalponorm, smith*). Figure 6.11(a) using the African *Timelines* data shows that the pairs are much closer and are also near the middle of the tree, whereas Figure 6.11(b) using the Indian *Decitalas* of 12 or 16 pulses, shows the pairs much farther apart, with *metricalponorm* near the bottom of the tree and *fitch, lhl*, and *smith* towards the top. This is the case for the other pairs as well; however, we note that the pairs (*oddity,tmavg*) and (*oddity,tmmax*) are farther apart using the African *Timelines* data and closer together using the Indian *Decitalas* data. However, the general trend is that one tree shows the pair closer while the other tree shows them far apart. To shed light on why this might be happening, consider Figure 7.6 showing boxplots of the raw complexity values of an outlier pair discussed above.

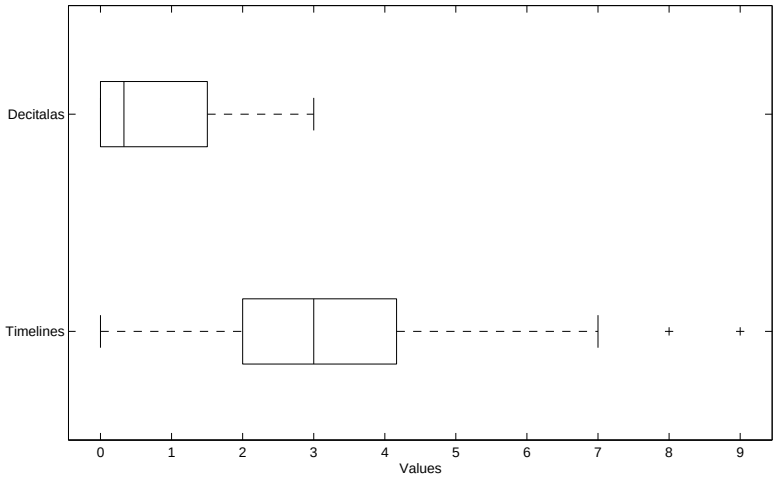
Notice from Figure 7.6 that in the comparison with the measure *metricalponorm*, the two distributions are similar in that they are both nearly symmetric. The main difference however, pertains to the range of complexity values, but this may be due to the nature of the data. Regarding the boxplots for the measure *lhl*, we can clearly see that there is another major difference other than the range of complexity values; i.e., the distribution of the Indian *Decitalas* is very skewed. In fact, this distribution is one-tailed. There are many possibilities for why this might be the case; however, because the *metricalponorm* is a normalized measure and avoids such skewness, perhaps the *lhl* measure can also benefit from a type of normalization.

Let us secondly discuss the outlier pairs for the African *Timelines* compared to the full data set of Indian *Decitalas*. In this comparison, it is clearly shown that the *offbeatness* measure is flagged as a very common measure among the outlier pairs. Quickly inspecting the phylogenetic trees in Figure 6.13 from Chapter 6 reveals that *offbeatness* is close to the middle of the tree using the African *Timelines* data set, but is then quite far for the Indian *Decitalas* data set.

We did not see this behavior from *offbeatness* when the Indian *Decitalas* data set was restricted to rhythms with 12 or 16 pulses. However, now that this constraint is dropped,



(a) Metrical Complexity (Onset and Pulse Normalized)



(b) Longuet-Higgins and Lee

Figure 7.6: *Comparison of a boxplot of the raw complexity values for the African* Timelines *of 12 or 16 pulses to that of the Indian* Decitalas *of 12 or 16 pulses. In (a) we have boxplots for* metricalponorm *and in (b) we have boxplots for* lhl*.*

the Indian *Decitalas* have a wide range of pulses. The mean number of pulses is 13.320 with a standard deviation of 6.748. Perhaps it is this varied range causing the *offbeatness* measure to show such deviant correlation compared to the more tame data set (with respect to pulses) of the African *Timelines*. Thus, normalization may be useful.

Lastly, consider the third set of outlier pairs from the comparison of the African *Timelines* to the North Indian *Decitalas*. The most notable aspect of the complexity measures contained in the outlier pairs is that they represent the top performers in the previous section pertaining to human-based complexity, which generally attained similar performance; i.e., they generally cluster together. Looking at Figure 6.13(a) using the African *Timelines* data, we can see that *fitch, keith, lhl, metricalonorm, pressing*, and *wnbd* cluster together in the middle of the tree. On the other hand, Figure 6.15(b) shows quite the opposite. The previous cluster of all seven measures now splits into two separate clusters: the first cluster consists of *keith, pressing*, and *wnbd*, whereas the second cluster consists of *fitch, lhl, metricalonorm*, and *smith*. Again, the reason for this behavior is perhaps due to the nature of the North Indian *Talas*. For this data the mean number of pulses is 13.357 with a standard deviation of 2.373. While this is not as dramatic as previously discussed, perhaps normalization for the number of pulses would make the results more consistent.

As a remark, a subtle difference seen between the two trees is the performance of *metricalonorm*, which implicitly is normalized for pulses and onsets. In the tree for the African *Timelines*, this measure is highly correlated to the *lhl*-type measures. In the tree for the North Indian *Talas*, the *metricalonorm* is still close to the *lhl*-type measures, but does reside further away; in fact the measure resides in its own sub-cluster with *ioi-l-mm* and *ioi-l-h*; two measures which also have implicit normalization for both pulses and onsets. The main conclusions, which may be drawn from comparing the performance of the complexity measures to data from different cultures is as follows.

1. Normalized complexity measures show a more symmetric distribution of complexity values across different cultural data as seen by the Metrical Complexity (Pulse and Onset Normalized) measure.
2. Even the top performing measures previously shown in the discussion of human complexity may benefit from a form of normalization to the number of pulses and perhaps even onsets.
3. The measures which have an implicit normalization are drawn closer together when the number of pulses is varied in the data.

7.3 Random Complexity

Robustness of the complexity measures was previously tested using cultural data of Africa and India. Here, random data is used to validate the tests between the cultures. The

random data is composed of two data sets: the first is a set of random rhythms with 12 pulses and the second is a set of random rhythms with 16 pulses. In the following sections, first the random rhythms with 12 pulses are compared to the random rhythms of 16 pulses. Second, the combined set of random rhythms including those with 12 or 16 pulses are compared to each cultural data set: the African *Timelines*, the Indian *Decitalas*, and the North Indian *Talas*.

7.3.1 Random vs. Random Robustness

Similar to the comparison of the cultural data with 12 pulses to that of 16 pulses, the results from using random rhythms are discussed here. Consider Table 7.4 which lists the outlier pairs from the results of the comparison between the random rhythms with 12 pulses compared to those with 16 pulses. The left column contains those complexity measures that when paired to each measure on the right, constitute an outlier.

Table 7.4: *Complexity measures which are considered unstable comparing the random rhythms with 12 pulses to random rhythms with 16 pulses*

Data Set	Complexity Measure Pair	
Random rhythms	dswap4	dswap, pressing, metricalonorm, metricalonormeuler, metricalponorm, wnbd
	ioi-l-h	ioi-g-h, ioi-g-mm
	wnbd2	fitch, keith, lhl, metricalonorm, metricalonormeuler, metricalponorm, metricalponormeuler, offbeatness, pressing, smith, wnbd

In Table 7.4, we see that *dswap4* and *wnbd2* are common among the outlier pairs. These results are quite similar to those from the comparison of African *Timelines* with 12 pulses to the *Timelines* with 16 pulses. However, here we also see that *ioi-l-h* is present in the outlier pairs. Going back to the phylogenetic trees in Figure 6.17, we can observe the changes in clustering regarding the complexity measures listed in the outlier pairs. First consider *dswap4*. In the random rhythms with 12 pulses (see Figure 6.17(a)), *dswap4* resides at the very bottom of the tree while the measures *dswap, pressing, metricalonorm, metricalonormeuler, metricalponorm*, and *wnbd*, are generally clustered in the upper half of the tree. For the random rhythms with 16 pulses (shown in Figure 6.17(b)), *dswap4* is located much closer to the cluster of the measures *pressing, metricalonorm, metricalonormeuler, metricalponorm*, and *wnbd*, near the center of the tree, and is tightly clustered to the *dswap* measure.

The *wnbd2* measure presents a similar scenario to that of the *dswap4*. However, in the tree for the random rhythms with 16 pulses, *wnbd2* is in fact inside the cluster composed of measures *fitch, keith, lhl, metricalonorm, metricalonormeuler, metricalponorm, metricalponormeuler, offbeatness, pressing, smith*, and *wnbd*, where it resides rather distant

from that same grouping of measures in the tree for the random rhythms with 12 pulses. Additionally, the random rhythm comparison results yield that *ioi-l-h* has significant change in correlation to *ioi-g-h* and *ioi-g-mm*. In fact, the trees in Figure 6.17 show that *ioi-l-h* is on the opposite side of the tree compared to *ioi-g-h* and *ioi-g-mm* (for the random rhythms of 16 pulses), but resides much closer in the main lower branch for the random rhythms of 12 pulses.

These findings support the previous conclusions that *wnbd* may benefit from a priori knowledge of which metrical divisions should be taken into account, as would the *dswap* measure. Additionally, it seems that the rhythms with 12 pulses admit higher correlation between local interonset interval histogram entropy measurements and global interonset interval histogram entropy measurements, as seen with *ioi-l-h*. Thus, we have the following conclusions.

1. The random rhythm comparison validates the intra-cultural comparison by showing that *dswap4* and *wnbd2* may not be appropriate divisions of the metrical structure for rhythms with 12 pulses.
2. Local IOI measures are more similar to global IOI measures when computing the entropy on the respective histograms from rhythms with 12 pulses compared to 16.

7.3.2 Random vs. Cultural Robustness

The random rhythms were compared to each of the cultural data sets to validate the performance of the measures. The goal was to test for those measures which acted unstably between each comparison. The random rhythm set is controlled to have a symmetric distribution of onsets. To visualize this, consider the following boxplot of the distribution of onsets for the combined data set of the random rhythms with 12 or 16 pulses, shown in Figure 7.7.

Figure 7.7 shows that the distribution for the onsets of the random rhythm is very symmetric and is quite ideal. Compared to the distributions of the number of onsets in the rhythms from the different cultures, the random rhythm distribution is quite different. This suggests that the random data sets are unrealistic in terms of rhythms that are actually found in different cultures, such as African or Indian. Thus, one may ask the question: *Do we want stable performance of complexity measures using an artificial, unrealistic data set, compared to realistic data sets from different cultures?* The answer is probably not.

Ideally, a rhythm complexity measure should accurately and robustly measure the complexity of rhythms created by humans, not those rhythms which have no human significance; i.e., are not found in any culture. The random rhythms here are not found in any of the cultural data sets used in this thesis, and so they are most likely not significant patterns according to humans. Therefore, we discuss the results of the performance of the complex-

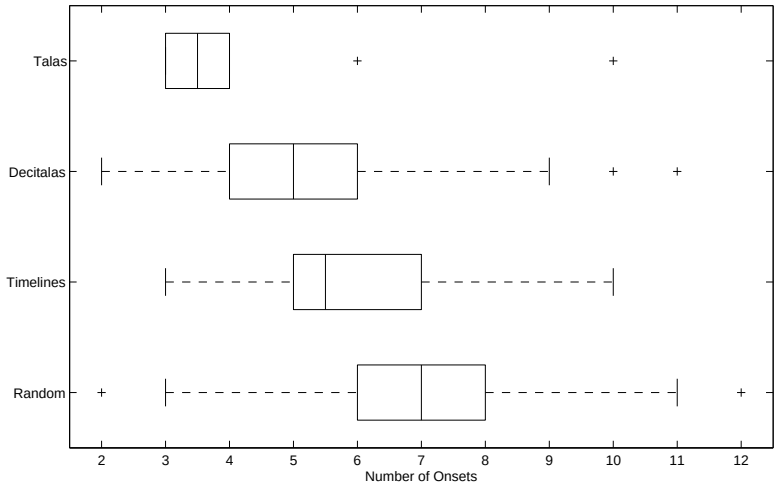


Figure 7.7: *Boxplot of the distribution of onsets from all the rhythms in the sets of random rhythms with 12 or 16 pulses, African* Timelines *with 12 or 16 pulses, Indian* Decitalas*, and North Indian* Talas*.*

ity measures on random rhythms compared to African and Indian rhythms, where we desire the measures to be unstable (outliers), indicating that perhaps such measures are more applicable to cultural data. Table 7.5, presents the outlier pairs from each comparison of the random rhythms to: African *Timelines*, Indian *Decitalas*, and North Indian *Talas*. Each set of outliers is separated by a horizontal line.

Table 7.5: *Complexity measures which are considered unstable comparing the random rhythms to the African* Timelines*, the Indian* Decitalas *below the first horizontal separator, and the North Indian* Talas *below the second horizontal separator.*

Data Set	Complexity Measure Pair	
African *Timelines*	hk	ioi-g-h, ioi-g-mm, offbeatness
	hrun	ceps, ioi-l-h, ioi-l-mm, ioi-l-sd
	ioi-g-h	hk, lz
	ioi-g-mm	hk, lz
	ioi-g-sd	lz, tmuavg, tmumax
	ioi-l-sd	hrun, lz
	lz	hrun, ioi-g-h, ioi-g-mm, ioi-g-sd, ioi-l-sd
	oddity	pressing, tmavg, tmmax
Indian *Decitalas*	offbeatness	dswap, fitch, hk, hrun, metricalponorm, tmuavg, tmumax
North Indian *Talas*	metricalonorm	keith, pressing, wnbd
	wnbd	fitch, lhl, metricalonorm, smith

A large number of different outlier pairs are listed in Table 7.5. Those which stand out are *ioi-g-h, ioi-g-mm, ioi-g-sd, ioi-l-sd*, and *oddity*. This suggests that perhaps complexity measures which depend on interonset intervals, and the rhythmic oddity property of the

Aka Pigmies, are more appropriate for studying rhythms of Africa, since they act as outliers between random and African rhythms.

The second set of outlier pairs yields familiar results. That is, *offbeatness* acts as an outlier regarding the Indian *Decitalas*. We previously concluded that high variance of the number of pulses in the *Decitalas* causes such a dramatic change for rhythms between different cultures. However, since the Off-Beatness measure is an outlier here, perhaps the measure may explain underlying mathematical properties inherent to Indian *Decitalas*.

In the third set of outlier pairs, we see a result similar to the African *Timelines* compared to the North Indian *Talas*. For the random rhythms, the measures that have correlated well with the human-based complexities are again shown to be outliers. Previously we concluded that for these measures to attain more robust behavior between cultures, normalization may be appropriate. Here, the results suggest that the type of normalization should be carefully chosen, as the measures may be useful for measuring the complexity of North Indian *Talas*.

From the discussion of the comparisons involving the random rhythm data set, we can make the following conclusions.

1. The measures which depend on interonset intervals or the rhythmic oddity property may be more appropriate for studying African derived rhythms.
2. The Off-Beatness measure shows drastic performance changes with random rhythms. This further supports the notion of normalization, but suggests that the measure may be appropriate for measuring the complexity of Indian *Decitalas*.
3. Measures which perform well compared with human-based measures are shown as outliers comparing random rhythms to North Indian *Talas*. This suggests that while normalization may be important, so is the type of normalization since the measures may retain properties appropriate for measuring the complexity of North Indian *Talas*.

7.4 Concluding Remarks

In the above discussion, we drew conclusions regarding human-based performance complexity, human-based perceptual complexity, and human-based metrical complexity. The most noteworthy result, is that of the 55 measures of rhythm complexity tested here, over half of them correlated well ($r_s^* > 0.500$) with the human-based measure of metrical complexity. This surprising result points to the type of rhythmic complexity the algorithms are most accurately measuring and thus, suggests two open problems.

Open Problem 1. *Design a complexity measure for rhythm performance complexity.*

Open Problem 2. *Design a complexity measure for rhythm perceptual complexity.*

Regarding *rhythm performance complexity*, the findings from the comparison of the complexity measures to human-based performance Complexity suggest that using the metrical hierarchy of Lerdahl and Jackendoff [99] with an alternate weighting scheme (such as that from Palmer and Krumhansl's work [114]) may yield the more accurate results. This is taking into consideration that the data will not be as controlled (i.e., artificial) as that used in Shmulevich and Povel's study [146]. Moreover, as a consequence of Fitch and Rosenfeld's work [57], the metrical hierarchy may not be the only technique needed to accurately measure performance complexity; i.e., using the information from the interonset interval histogram as seen from Toussaint [172, 173], and also incorporating the notion of musical elaboration from Tanguiane [157], may yield improved results.

As for *rhythm perceptual complexity*, the findings from the comparison of the complexity measures to human-based perceptual complexity suggest that the metrical hierarchy may also yield accurate results. However, since the nature of the data was not quite adequate to solidify the findings of the best weighting scheme to use for the hierarchy, improvements may possibly be found by using a scheme from Palmer and Krumhansl [114] or even Euler's scheme from Gonzalez [69]. Moreover, perhaps the Rhythmic Oddity property [4], which was then generalized to a rhythm complexity measure by Toussaint [169], may be useful in measuring perceptual complexity. However, more psychological validation is necessary, which in fact points to an open problem in the psychological domain.

Open Problem 3. *Design a psychological experiment for measuring human-based performance and perceptual complexity using realistic rhythm data.*

In addition to the design of new algorithms, improvements on current ones may also be made in an attempt to more accurately measure the general notion of complexity. Especially, when using *realistic* rhythm data; i.e, rhythms representative of different cultures. From the comparisons of the performance of the complexity measures on such data, it is clear that there exists a gap when handling the inherent complexities of such realistic data; e.g., a different number of pulses.

By making comparisons among the African *Timelines*, Indian *Decitalas*, and the random rhythms, the results suggest a need for a way to appropriately normalize. Moreover, the need for a generalization and improvement of certain measures was discussed. One example of a measure which would benefit from generalization is Keith's Complexity measure, as it is only appropriate for rhythms where the number of pulses is a power of two.

Another example for improving an algorithm may be to use a less-strict scheme for the Off-Beatness measure. For example, instead of strictly labeling each pulse as *on-beat* or *off-beat*, perhaps a pulse can be off-the-beat by some weight, where the lowest weight would constitute being on-the-beat. One way to accomplish this would be that upon inscription

of the polygons which evenly sub-divide the number of pulses, one can keep track of the number of vertices which touch a pulse. Since all polygons are inscribed starting at the zeroth pulse, that pulse would have the lowest off-beat weight. However, in the case of 16 pulses, pulse three has a vertex from a square and an octagon, so that pulse would have a higher off-beat weight. Using this weighting scheme may improve the measure by giving a larger spread to the range of complexity values. The WNBD measure and the Directed Swap measure may also benefit from extracting more information about the rhythms being measured. This would allow for a more robust beat-detection, which may improve results.

Specific to the random rhythm and cultural data results, the type of normalization is also a large concern. Since measures like Rhythmic Oddity, the IOI-based measures, the Off-Beatness measure, and those which correlate well with human-based judgements are shown to be outliers regarding random versus cultural data sets, there may be useful properties we would like to preserve in the measures, and not loose when applying a form of normalization. Therefore, there is a necessity for the study of a larger open problem.

Open Problem 4. *Design a method for the appropriate normalization of rhythm complexity values when there are few constraints on the rhythm input; i.e., the number of pulses and/or onsets may be varied.*

In conclusion, each open problem presented is the future work of this thesis, as is the continued validation of newly discovered rhythm complexity measures. Current techniques have been analyzed for improvements and a general short-coming has been posed in terms of an open problem. Once such problems have been solved, and appropriate algorithms for the complexity of musical rhythm have been discovered, then perhaps one final open problem may be studied.

Open Problem 5. *Which culture has the most complex musical rhythms?*

Such a problem suggests that the complexity of musical rhythm requires careful study. This thesis has explored a variety of ways in which the complexity of musical rhythm can be measured. It is the hope that such methods seen here may be refined to handle some of the problems encountered with realistic rhythm data, and then incorporated into computer-based tools for studying rhythm. Such tools are present in the field of Music Information Retrieval as feature extraction tools. This thesis is not alone in considering complexity to be a feature of music, and so this lends itself well to the MIR domain. Perhaps if rhythm complexity is incorporated into MIR tools, then those in the field of computational ethnomusicology and musicology alike may find such computer assistance beneficial in analyzing the complexity of musical rhythm, and maybe one day determine the open problem posed here of which culture has the most complex musical rhythms.

References

[1] K. Agawu. Structural analysis of cultural analysis? Competing perspectives on the 'standard pattern' of West African rhythm. *Journal of the American Musicological Soceity*, 59(1):1–46, 2006.

[2] W. Anku. Circles and time: a theory of structural organization of rhythm in African music. *Music Theory Online*, 6(1), Jan 2000.

[3] A. V. Arkhangel'Skii and L. S. Pontryagin. *General Topology I: Basic Concepts and Constructions Dimension Theory (Encyclopaedia of Mathematical Sciences)*. Springer, 1990.

[4] S. Arom. *African Polyphony and Polyrhythm*. Cambridge University Press, 2004.

[5] M. Babbitt. The use of computers in musicological research. *Perspectives of New Music*, 3(2):74–83, Spring–Summer 1965.

[6] D. Bainbridge, C. G. Nevill-Manning, I. H. Witten, L. A. Smith, and R. J. McNab. Towards a digital library of popular music. In *Proceedings of the Fourth ACM International Conference on Digital Libraries*, pages 161–169. Association for Computing Machinery, Aug 1999.

[7] D. W. Barber. *The Music Lover's Quotation Book, Third Edition*. Sound And Vision Publishing, Ltd., 2003.

[8] A. Binder and B. R. Wolin. Information models and their uses. *Psychometrika*, 29(1):29–54, 1964.

[9] S. Block and J. Douthett. Vector products and intervallic weighting. *Journal of Music Theory*, 38(1):21–41, 1994.

[10] A. Bookstein. Fuzzy hamming distance: a new dissimilarity measure. In *Combinatorial Pattern Matching:* 12^{th} *Annual Symposium*, volume 2089 of *Lecture Notes in Computer Science*, pages 86–97. Springer Berlin/Heidelberg, Jul 2001.

[11] E. A. Bowles. Musicke's handmaiden: or technology in the service of the arts. In H. B. Lincoln, editor, *The Computer and Music*, pages 3–20. Cornell University Press, 1970. Chapter I.

[12] B. H. Bronson. Mechanical help in the study of folk song. *The Journal of American Folklore*, 62(244):81–86, Apr–Jun 1949.

[13] D. Byrd and T. Crawford. Problems of music information retrieval in the real word. *Information Processing and Management*, 38:249–272, 2002.

[14] J. H. Camin and R. R. Sokal. A method for deducing branching sequences in phylogeny. *Evolution*, 19(3):311–326, Sep 1965.

[15] J. B. Carroll. The nature of the data, or how to choose a correlation coefficient. *Psychometrika*, 26(4):347–372, Dec 1961.

[16] L. L. Cavalli-Sforza and A. W. F. Edwards. Phylogenetic analysis: models and estimation procedures. *Evolution*, 21:550–570, Sep 1967.

[17] M. Chemlillier and C. Truchet. Computation of words satisfying the "rhythmic oddity property" (after Simha Arom's work). *Information Processing Letters*, 86:255–261, 2003.

[18] J. C. C. Chen and A. L. P. Chen. Query by rhythm: and approach for song retrieval in music databases. In *In proceedings of the International Workshop on Research Issues in Data Engineering*, pages 139–146, 1998.

[19] P. Y. Chen and P. M. Popovich. *Correlation: Parametric and Nonparametric Measures*. Quantitative Applications in the Social Sciences. Sage Publications, 2002.

[20] M. Clayton. *Time in Indian Music*. Oxford Monographs on Music. Oxford University Press, 2000.

[21] A. D. Cliff, P. Haggett, M. R. Smallman-Raynor, D. F. Stroup, and G. D. Williamson. The application of multidimensional scaling methods to epidemiological data. *Statistical Methods in Medical Research*, 4:102–123, 1995.

[22] J. Clough. Aspects of diatonic sets. *Journal of Music Theory*, 23(1):45–61, 1979.

[23] J. Clough and G. Myerson. Variety and multiplicity in diatonic systems. *Journal of Music Theory*, 29:249–270, 1985.

[24] J. Cohen and P. Cohen. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, Second Edition*. Lawrence Erlbaum Associates, 1983.

[25] S. Cohen. *Finding color and shape patterns in images*. PhD thesis, Stanford University, 1999.

[26] J. Colannino, M. Damian, F. Hurtado, J. Iacono, H. Meijer, S. Ramaswami, and G. T. Toussaint. An O(n log n)-time algorithm for the restriction scaffold assignment problem. *Journal of Computational Biology*, 13(4):979–989, 2006.

[27] J. Colannino, F. Gómez, and G. T. Toussaint. Analysis of emergent beat-class sets in Steve Reich's clapping music and the Yoruba bell timeline. *Perspectives of New Music*, 2008. Accepted for publication.

[28] J. Colannino and G. T. Toussaint. An algorithm for computing the restriction scaffold assignment problem in computational biology. *Information Processing Letters*, 95(4):466–471, Aug 2005.

[29] G. W. Cooper and L. B. Meyer. *The Rhythmic Structure of Music*. The University of Chicago Press, 1960.

[30] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein. *Introduction to Algorithms, Second Edition*. The MIT Press, 2001.

[31] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.

[32] H. S. M. Coxeter. Music and mathematics. *The Canadian Music Journal*, VI:13–24, 1962.

[33] E. J. Coyle and I. Shmulevich. A system for machine recognition of music patterns. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 6, pages 3597–3600, May 1998.

[34] L. Cronbach. Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3):297–334, 1951.

[35] I. Cross. Music and biocultural evolution. In M. Clayton, T. Herbert, and R. Middleton, editors, *The Cultural Study of Music: A Critical Introduction*, pages 19–30. Routledge, 2003.

[36] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order, Second Edition*.

Cambridge University Press, 2005.

[37] N. de Bruijin. Sorting by means of swapping. *Discrete Mathematics*, 9:333–339, 1974.

[38] M. J. L. de Hoon, S. Imoto, J. Nolan, and S. Miyano. Open source clustering software. *Bioinformatics*, 20(9):1453–1454, 2004.

[39] L. Devroye. Random number generation (Luc Devroye). http://cg.scs.carleton.ca/~luc/lecuyer.c. Accessed: 29 May 2008.

[40] J. M. Díaz-Báñez, G. Farigu, F. Gómez, D. Rappaport, and G. T. Toussaint. El compás flamenco: A phylogenetic analysis. In *BRIDGES: Mathematical Connections in Art, Music and Science*, Jul 2004.

[41] E. J. Dietz. Permutation tests for association between two distance matrices. *Systematic Zoology*, 32(1):21–26, 1983.

[42] J. S. Downie. The scientific evaluation of music information retrieval systems: foundations and future. *Computer Music Journal*, 28(2):12–23, Summer 2004.

[43] A. Dress, D. Huson, and V. Moulton. Analyzing and visualizing sequence and distance data using SplitsTree. *Discrete Applied Mathematics*, 71:95–109, 1996.

[44] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification, Second Edition*. Wiley-Interscience, 2000.

[45] A. W. F. Edwards and L. L. Cavalli-Sforza. *The reconstruction of evolution*, volume 6, pages 67–76. Systematics Assoc. Publ., 1964.

[46] A. W. F. Edwards and L. L. Cavalli-Sforza. A method for cluster analysis. *Biometrics*, 21:362–375, 1965.

[47] T. Eeerola, T. Himberg, P. Toivainen, and J. Louhivouri. Perceived complexity of western and African folk melodies by western and African listeners. *Pscyhology of Music*, 34(3):337–371, 2006.

[48] P. Essens. Structuring temporal sequences: Comparison of models and factors of complexity. *Perception and Psychophysics*, 57(4):519–532, 1995.

[49] B. S. Everitt, S. Landau, and M. Leese. *Cluster Analysis, Fourth Edition*. Oxford University Press, Inc., 2001.

[50] J. Felsenstein. Maximum likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Zoology*, 22(3):240–249, Sep 1973.

[51] J. Felsenstein. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, 17:368–376, 1981.

[52] J. Felsenstein. Phylogeny programs. http://evolution.genetics.washington.edu/phylip/software.html, 2008.

[53] R. Fiebrink and I. Fujinaga. Feature selection pitfalls and music classification. In *Proceedings of the International Conference on Music Information Retrieval*, pages 340–340, 2006.

[54] W. M. Fitch and E. Margoliash. Construction of phylogenetic trees. *Science, New Series*, 155(3760):279–284, Jan 1967.

[55] W. T. Fitch. Personal communication, 2007.

[56] W. T. Fitch and A. Rosenfeld. Implementation of the Longuet-Higgins and Lee algorithm in Python. http://www.st-andrews.ac.uk/~wtsf/rhythmFiles/SyncopationIndex.py, 2007.

[57] W. T. Fitch and A. J. Rosenfeld. Perception and production of syncopated rhythms. *Music Perception*, 25(1):43–58, 2007.

[58] A. Forte. The domain and relations of set-complex theory. *Journal of Music Theory*, 9(1):173–180, Spring 1965.

[59] Free Software Foundation. GCC, the GNU compiler collection. http://gcc.gnu.org/, 2007.

[60] M. Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200):675–701, Dec 1937.

[61] M. Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance (a correction). *Journal of the American Statistical Association*, 34(205):109, Mar 1939.

[62] I. Fujinaga and D. McEnnis. On-demand metadata extraction network (OMEN). In *Proceedings of the Joint Conference on Digital Libraries*, pages 346–346, 2006.

[63] I. Fujinaga, S. Moore, and D. S. Sullivan. Implementation of exemplar-based learning model for music cognition. In S. W. Yi, editor, *Music, Mind, and Science*, pages 69–81. Seoul National University Press, 1999.

[64] W. R. Garner. *Uncertainty and Structure as Psychological Concepts*. John Wiley & Sons, Inc., 1962.

[65] O. Gascuel. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution*, 14(7):685–695, 1997.

[66] Gentoo Linux Foundation, Inc. Gentoo Linux. http://www.gentoo.org/, 2008.

[67] F. Gómez, A. Melvin, D. Rappaport, and G. Toussaint. Mathematical measures of syncopation. In *BRIDGES: Mathematical Connections in Art, Music and Science*, pages 73–84, Jul 2005.

[68] F. Gómez, E. Thul, and G. Toussaint. An experimental comparison of formal measures of rhythmic syncopation. In *Proceedings of the International Computer Music Conference*, pages 101–104, Aug 2007.

[69] F. J. S. González. A numerical theory of rhythm applied to oriental music analysis. http://www.uam.es/personal_pdi/filoyletras/jsango/a_numerical_theory_of_rhythm.html. Accessed: 2008 May.

[70] J. K. Haak. Clapping music–a combinatorial problem. *The College Mathematics Journal*, 22(3):224–227, May 1991.

[71] J. K. Haak. The mathematics of Steve Reich's clapping music. In *Proceedings of BRIDGES: Mathematical Connections in Art, Music, and Science*, pages 87–92, 1998.

[72] J. Heringa. Phylogenetic tree (phylogenetic reconstruction, phylogeny, phylogeny reconstruction). In *Dictionary of Bioinformatics and Computational Biology*. John Wiley & Sons, Inc., Oct 2004.

[73] P. Hogeweg and B. Hesper. The alignment of sets of sequences and the construction

of phyletic trees: an integrated method. *Journal of Molecular Evolution*, 20:175–186, 1984.

[74] M. Hollander and D. A. Wolfe. *Nonparametric Statistical Methods, Second Edition*. John Wiley & Sons, Inc., 1999.

[75] H. Hotelling and M. R. Pabst. Rank correlation and tests of significance involving no assumption of normality. *The Annals of Mathematical Statistics*, 7(1):29–43, Mar 1936.

[76] D. Huron. Lecture 4. What is a musical feature? Forte's analysis of Brahms opus 51, no. 1 revisited. http://musiccog.ohio-state.edu/Music220/Bloch.lectures/4.Analysis.html, 1999. The 1999 Ernest Bloch Lectures.

[77] D. Huron and A. Ommen. An empirical study of syncopation in American popular music, 1890–1939. *Music Theory Spectrum*, 28:211–231, 2006.

[78] D. H. Huson. SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics*, 14(1):68–73, 1998.

[79] D. H. Huson and D. Bryant. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2):254–267, 2006.

[80] J. W. Jaromczyk and G. T. Toussaint. Relative neighborhood graphs and their relatives. *Proc. IEEE*, 40(9):1502–1517, Sep 1992.

[81] R. S. Johnson. *Messiaen*. University of California Press, 1975.

[82] A. M. Jones. African rhythm. *Africa*, 24(1):26–47, Jan 1954.

[83] W. Just. Complexity issues in bioinformatics. http://www.math.ohiou.edu/~just/TALKS/Complexitytalk/, Apr 2002. Presentation.

[84] M. Kassler. Toward music information retrieval. *Perspectives of New Music*, 4(2):59–67, Spring–Summer 1966.

[85] I. J. Katz. Flamenco. In L. Macy, editor, *Grove Music Online*. Oxford University Press, 2008. Accessed: 07 June 2008.

[86] M. Keith. *From Polychords to Pólya: Adventures in Musical Combinatorics*. Vinculum Press, 1991.

[87] M. Kendall and J. D. Gibbons. *Rank Correlation Methods, Fifth Edition*. Oxford University Press, 1990.

[88] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

[89] B. W. Kernighan and D. M. Ritchie. *The C Programming Language, Second Edition*. Prentice Hall, 1988.

[90] J. Koetting. Analysis and notation of west African drum ensemble music. *Selected Reports in Ethnomusicology*, 1(3):115–146, 1970.

[91] W. H. Kruskal. Ordinal measures of association. *Journal of American Statistical Association*, 53(284):814–861, Dec 1958.

[92] P. L'Ecuyer. Random number generator C code. http://cg.scs.carleton.ca/~luc/lecuyer.c. Accessed: 29 May 2008.

[93] P. L'Ecuyer. Uniform random number generation. *Annals of Operations Research*, 53:77–120, 1994.

[94] P. L'Ecuyer. Uniform random number generators. In D. J. Medeiros, E. F. Watson, J. S. Carson, and M. S. Manivannan, editors, *Proceedings of the 1998 Winter Simulation Conference*, Winter 1998.

[95] P. L'Ecuyer and P. Hellekalek. Random number generators: selection criteria and testing. In P. Hellekalek and G. Larcher, editors, *Random and Quasi-Random Point Sets*, volume 138 of *Lecture Notes in Statistics*, pages 223–265. Springer, 1998.

[96] L. P. Lefkovitch. A nonparametric method for comparing dissimilarity matrices, a general measures of biogeographical distance, and their application. *The American Naturalist*, 123(4):484–499, Apr 1984.

[97] L. P. Lefkovitch. Further nonparametric tests for comparing dissimilarity matrices ased on the relative neighborhood graph. *Mathematical Biosciences*, 73:71–88, 1985.

[98] A. Lempel and J. Ziv. On the complexity of finite sequences. *IEEE Transactions on Information Theory*, IT-22(1):75–81, 1976.

[99] F. Lerdahl and R. Jackendoff. *A Generative Theory of Tonal Music*. MIT Press, 1983.

[100] D. Lewin. Forte's interval vector, My interval function, and Regener's common-note function. *Journal of Music Theory*, 21(2):194–237, Autumn 1977.

[101] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and its Applications*. Springer, 1997.

[102] Y. Liu and G. T. Toussaint. A tessellation-transformation method for categorizing geometric textile design patterns. In *Design Conference 2008: Second International Conference on Design Principles and Practices*, Jan 2008.

[103] S. Lloyd. Measures of complexity: a nonexhaustive list. *IEEE Control Systems Magazine*, 21(4):7–8, 2001.

[104] S. Lloyd. *Programming the Universe*. Alfred A. Knopf, 2006.

[105] J. London. Rhythm. In L. Macy, editor, *Grove Music Online*. Oxford University Press, 2008. Accessed: 20 March 2008.

[106] H. Longuet-Higgins and C. Lee. The rhythmic interpretation of monophonic music. *Music Perception*, 1(4):424–441, 1984.

[107] B. J. McCartin. Prelude to musical geometry. *The College Mathematical Journal*, 29(5):354–370, 1998.

[108] C. McKay and I. Fujinaga. Style-independent computer-assisted exploratory analysis of large music collections. *Journal of Interdisciplinary Music Studies*, 1(1):63–85, Spring 2007.

[109] O. Messiaen. *Technique de Mon Langage Musical*. Leduc Publications, 2000. Original publication c. 1944.

[110] B. Mont-Reynaud and M. Goldstein. On finding rhythmic patterns in musical lines. In *Proceedings of the International Computer Music Conference*, pages 391–397, 1985.

[111] J. L. Myers and A. D. Well. *Research Design and Statistical Analysis, Second Editon*. Lawrence Erlbaum Associates, 2003.

[112] P. Nauert. Theory and practice in Porgy and Bess: the Gershwin-Schillinger connec-

tion. *The Musical Quarterly*, 78(1):9–33, Spring 1994.

[113] M. J. O'Brien, J. Darwent, and R. L. Lyman. Cladistics is useful for reconstructing archaeological phylogenies: Palaeoindian points from the southeastern United States. *Journal of Archaeological Science*, 28:1115–1136, 2001.

[114] C. Palmer and C. L. Krumhansl. Mental representations for musical meter. *Journal of Experimental Psychology*, 16(4):728–741, 1990.

[115] R. Parncutt. Systematic musicology and the history and future of western musical scholarship. *Journal of Interdisciplinary Music Studies*, 1(1):1–32, Spring 2007. Article No. 071101.

[116] K. Pearson. Mathematical contributions to the theory of evolution. XVI. on further methods of determining correlation. In *Drapers' Co. Research Memoirs*. Cambridge University Press, 1907.

[117] K. Pearson. On an extension of the method of correlation by grades or ranks. *Biometrika*, 10(2/3):416–418, 1914.

[118] E. Pekalsak, A. Harol, R. P. W. Duin, B. Spillmann, and H. Bunke. *Non-Euclidean or Non-Metric Measures can be Informative*, volume 4109 of *Lecture Notes in Computer Science*, pages 871–880. Springer Berlin/Heidelberg, 2006.

[119] D. Povel and P. Essens. Perception of temporal patterns. *Music Perception*, 2:411–440, 1985.

[120] D.-J. Povel. Internal representtion of simple temporal patterns. *Journal of Experimental Psychology*, 7(1):3–18, 1981.

[121] D.-J. Povel. A theoretical framework for rhythmic perception. *Psychological Research*, 45:315–337, 1984.

[122] H. S. Powers. Reviewed work(s): Saṅgīta-Ratnākara of Ṡarṅgadeva by R. K. Shringy. *Ethnomusicology*, 28(2):352–355, May 1984.

[123] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing, Second Edition*. Cambridge University Press, 1992.

[124] J. Pressing. Cognitive isomorphisms in world music: West Africa, the Balkans, Thailand and Western tonality. *Studies in Music*, 17:38–61, 1983.

[125] J. Pressing. Cognitive complexity and the structure of musical patterns. http://www.psych.unimelb.edu.au/staff/jp/cog-music.pdf, 1999.

[126] J. Pressing and P. Lawrence. Transcribe: a comprehensive autotranscription program. In *Proceedings of the International Computer Music Conference*, pages 343–345, 1993.

[127] Python Software Foundation. The Python programming language. http://www.python.org/, 2008. Version 2.5.1.

[128] J. Rahn. Asymmetrical ostinatos in Sub-Saharan music: time, pitch, and cycles reconsidered. *In Theory Only*, 9(7):23–37, 1987.

[129] J. Rahn. Turning the analysis around: Africa-derived rhythms and Europe-derived music theory. *Black Music Research Journal*, 16(1):71–89, 1996.

[130] D. M. Randel, editor. Syncopation. In *The Harvard Dictionary of Music, Fourth Edition*, Harvard University Press Reference Library, pages 861–862. The Belknap

Press of Harvard University, 2003.

[131] E. Regener. On Allen Forte's theory of chords. *Perspectives of New Music*, 13(1):191–212, 1974.

[132] S. Reich. *Writings about Music*. New York University Press, 1974.

[133] C. Riedweg. *Pythagoras: His Life, Teaching, and Influence*. Cornell University Press, 2005. Translated by Steven Rendall.

[134] T. D. Robison. IML-MIR: a data processing system for the analysis of music. In H. von Harald Heckmann, editor, *Elektronische Datenverarbeitung in der Musikwissenschaft*, pages 103–135. Gustav Bosse Verlag, 1967.

[135] J. L. Rue. On style analysis. *Journal of Music Theory*, 6(1):91–107, Spring 1962.

[136] C. Sachs. *Rhythm and Tempo: A Study in Music History*. W. W. Nortan & Company, Inc., 1953.

[137] N. Saitou and M. Nei. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4(4):406–425, Jul 1987.

[138] J. Schillinger. *The Schillinger System of Musical Composition*. Carl Fischer, 1946.

[139] J. Schillinger. *The Mathematical Basis of the Arts*. Philosophical Library, 1948.

[140] J. Schillinger. *Encyclopedia of Rhythms*. Da Capo Press, 1976.

[141] T. D. Schneider. Information theory primer. http://www.lecb.ncifcrf.giv/~toms/paper/primer/, Jun 2000. National Cancer Institute, Frederick Cancer Research and Development Center, Laboratory of Experimental and Computational Biology.

[142] C. Seeger. An instantaneous music notator. *Journal of the International Folk Music Council*, 3:103–106, 1951.

[143] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423; 623–656, Jul; Oct 1948.

[144] I. Shmulevich and D.-J. Povel. Rhythm complexity measures for music pattern recognition. In *IEEE 2nd Workshop on Multimedia Signal Processing*, pages 167–172, 1998.

[145] I. Shmulevich and D.-J. Povel. Complexity measures of musical rhythms. In P. Desain and L. Windsor, editors, *Rhythm Perception and Production*, pages 239–244. Swets & Zeitlinger, 2000.

[146] I. Shmulevich and D.-J. Povel. Measures of temporal pattern complexity. *Journal of New Music Research*, 29(1):61–69, 2000.

[147] I. Shmulevich, O. Yli-Harja, E. Coyle, D.-J. Povel, and K. Lemstrőm. Perceptual issues in music pattern recognition: complexity of rhythm and key finding. *Computers and the Humanities*, 35:23–35, 2001.

[148] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Monographs on Statistics and Applied Probability. Chapman and Hall, 1986.

[149] H. Situngkir. Towards complexity studies of Indonesian songs. Technical Report WP2007, Bandung Fe Institute, Aug 2007. BFI Working Paper Series: http://ssrn.com/abstract=1007509.

[150] L. Smith and H. Honing. Evaluating and extending computational models of rhythmic syncopation in music. In *Proceedings of the International Computer Music*

Conference, pages 688–691, 2006.

[151] L. M. Smith. Personal communication, Apr 2007.

[152] C. Spearman. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101, 1904.

[153] C. Spearman. A footrule for measuring correlation. *British Journal of Psychology*, 2:89–108, 1906.

[154] R. H. Stetson. A motor theory of rhythm and discrete succession. I. *Psychological Review*, 12(4):250–270, Jul 1905.

[155] G. Strangman. Python modules *stats* and *pstats*. http://www.nmr.mgh.harvard.edu/Neural_Systems_Group/strang/python.html, Feb 2002.

[156] J. S. Tanaka and G. J. Huba. A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, 38:197–201, 1985.

[157] A. S. Tanguiane. *Artificial Perception and Music Recognition*. Springer-Verlag, 1993.

[158] A. S. Tanguiane. A principle of correlativity of perception and its application to music recognition. *Music Perception*, 11(4):465–502, 1994.

[159] P. Taslakain and G. T. Toussaint. Geometric properties of musical rhythms. In *Proceedings of the* 16^{th} *Fall Workshop on Computational and Combinatorial Geometry*, Nov 2006.

[160] J. Tehrani and M. Collard. Investigating cultural evolution through biological phylogenetic analysis of Turkmen textiles. *Journal of Anthropological Archaeology*, 21:443–463, 2002.

[161] D. Temperley. Syncopation in rock: a perceptual perspective. *Popular Music*, 18(1):19–40, Jan 1999.

[162] The GNOME Project and PyGTK Team. PyGTK: GTK+ for Python. http://www.pygtk.org/, 2008. Version 2.10.6.

[163] The MathWorks™. MATLAB®: the language of technical computing. Commercial Software, Aug 2006. Version: 7.3.0.298 (R2006b).

[164] E. Thul and G. T. Toussaint. Analysis of musical rhythm complexity measures in a cultural context. In B. C. Desai, editor, *Proceedings of the Canadian Conference on Computer Science and Software Engineering*, pages 1–9, May 2008. Invited paper.

[165] E. Thul and G. T. Toussaint. A comparative phylogenetic analysis of African timelines and North Indian talas. In *Proceedings of BRIDGES: Mathematical Connections in Art, Music, and Science*, Jul 2008. In press.

[166] G. T. Toussaint. A mathematical analysis of African, Brazilian, and Cuban clave rhythms. In *BRIDGES: Mathematical Connections in Art, Music and Science*, pages 157–168, Jul 2002.

[167] G. T. Toussaint. A mathematical analysis of African, Brazilian, and Cuban clave rhythms. http://cgm.cs.mcgill.ca/~godfried/publications/clave.pdf,, May 2002. Extended.

[168] G. T. Toussaint. Classification of phylogenetic analysis of African ternery rhythm timelines. In *BRIDGES: Mathematical Connections in Art, Music and Science*, pages

23–27, Jul 2003.

[169] G. T. Toussaint. Classification of phylogenetic analysis of African ternery rhythm timelines. http://cgm.cs.mcgill.ca/~godfried/publications/ternary.pdf, Aug 2003. Extended.

[170] G. T. Toussaint. The Euclidean algorithm generates traditional musical rhythms. In *Proceedings of BRIDGES: Mathematical Connections in Art, Music, and Science*, pages 47–56, Jul 2005.

[171] G. T. Toussaint. The Euclidean algorithm generates traditional musical rhythms. http://cgm.cs.mcgill.ca/~godfried/publications/banff.pdf, 2005. Extended.

[172] G. T. Toussaint. The geometry of musical rhythm. In J. Akiyama, M. Kano, and X. Tan, editors, *Proceedings of the Japan Conference on Discrete and Computational Geometry*, volume 3742 of *Lecture Notes in Computer Science*, pages 198–212. Springer Berlin/Heidelberg, 2005.

[173] G. T. Toussaint. A comparison of rhythmic similarity measures. *FORMA*, 21(2):129–149, 2006.

[174] G. T. Toussaint. Personal communication, Mar 2007.

[175] J. W. Tukey. *Exploratory Data Analysis*. Addison-Wesley, 1977.

[176] R. Typke, P. Giannopoulos, R. C. Veltkamp, F. Wiering, and R. van Oostrum. Using transportation distances for measuring melodic similarity. In *International Conference on Music Information Retrieval*, pages 107–114, Oct 2003.

[177] G. Tzanetakis, A. Kapur, W. A. Schloss, and M. Wright. Computational ethnomusicology. *Journal of Interdisciplinary Music Studies*, 1(2):1–24, Fall 2007.

[178] University of Tübingen. Splitstree.org. http://www.splitstree.org/, 2007.

[179] P. C. Vitz. Information, run structure and binary pattern complexity. *Perception and Psychophysics*, 3(4A):275–280, 1968.

[180] P. C. Vitz and T. C. Todd. A coded element model of the perceptual processing of sequential stimuli. *Psycological Review*, 75(6):433–449, Sep 1969.

[181] E. M. von Hornbostel. African negro music. *Africa: Journal of the International African Institute*, 1(1):30–62, Jan 1928.

[182] H. E. Weaver. Syncopation: a study of musical rhythms. *The Journal of General Psychology*, 20:409–429, 1939.

[183] E. W. Weisstein. "Polygon Diagonal", from MathWorld–a Wolfram web resource. http://mathworld.wolfram.com/PolygonDiagonal.html. Accessed: 27 May 2008.

[184] R. C. Winkworth, D. Bryant, P. J. Lockhart, D. Havell, and V. Moulton. Biogeographic interpretation of splits graphs: least squares optimization of branch lengths. *Systematic Biology*, 54(1):56–65, 2005.

[185] M. Yeston. *The Stratification of Musical Rhythm*. Yale University Press, 1976.

[186] J. Ziv and A. Lempel. A universal algorithm for sequential data compression. *IEEE Transactions on Information Theory*, IT-23(3):337–343, 1977.

[187] J. Ziv and A. Lempel. Compressions of individual sequences via variable-rate encoding. *IEEE Transactions on Information Theory*, IT-24(5):530–536, 1978.